Joint Ex. 1

(JCCX33 - Alan Manning, Monopsony in Motion (Princeton University Press, 2003))

PART 2

the behavior of labor markets. So it is the simple model that we use in later chapters.

Appendix 5

Proof of Proposition 5.1

Suppose that $w(t)$ is not equal to p. We show that moving $w(t)$ to p with an appropriate off-setting adjustment to B can never reduce profits and will increase them as long as there is some effect on the separation rate. Suppose that the initial contract is $\{w(t), B\}$. Now consider replacing it by a contract in which $w(t) = p$ for all t and B is given by B_1 where B_1 satisfies

$$B_1 = B + \frac{\int_0^\infty [p - w(t)]N(t)dt}{R} \tag{5.11}$$

where R is the level of recruits in the old contract, and $N(t)$ is the level of employment of workers with tenure t. As the new contract only makes profits from the entrance fee, profits from the new contract will be R_1B_1 where R_1 is the level of recruits in the new contract so that, using (5.11), we have

$$\Pi_1 = R_1B_1 = \frac{R_1}{R}\left[RB + \int_0^\infty [p - w(t)]N(t)dt\right] = \frac{R_1}{R}\Pi \tag{5.12}$$

where Π is the level of profits on the old contract. So, profits cannot fall with the new contract if $R_1 \geq R$ and will rise if this is a strict inequality. Whether recruits rise depends on whether $V_1(0) - B_1$ is larger than $V(0) - B$.

To show that recruits cannot fall, consider the following argument. By integrating (5.1). we can write the value of the job under the old contract as

$$V(t) = \int_t^\infty \left[w(\tau) + \delta_u V^u + \lambda \int_{V(\tau)} VdF(V)\right]$$

$$\times \exp\left(-\int_t^\tau [\delta + \lambda[1 - F(V(\tau'))]]d\tau'\right)d\tau \tag{5.13}$$

Now suppose that the quit decision of workers remains the same in the old contract as in the new contract (it will actually be lower as wages are raised to p but we will return to that). Then the value of the job under the new contract will be given by

$$V_1(t) = \int_t^\infty \left[p + \delta_u V^u + \lambda \int_{V(\tau)} V dF(V)\right]$$

$$\times \exp\left[-\int_t^\tau \delta + \lambda[1 - F(V(\tau'))]d\tau'\right]d\tau \qquad (5.14)$$

Hence the difference in the value of the two contracts to a new recruit can be written as

$$[V_1(0) - B_1] - [V(0) - B]$$

$$= \int_0^\infty [p - w(\tau)] \exp\left[-\int_0^\tau [\delta + \lambda[1 - F(V(\tau'))]]d\tau'\right]d\tau - B_1 + B \quad (5.15)$$

Now, from (5.2) we can write $N(t)$ as

$$N(\tau) = R \exp\left[-\int_0^\tau [\delta + \lambda[1 - F(V(\tau'))]]d\tau'\right] \qquad (5.16)$$

so that (5.15) can be written as

$$[V_1(0) - B_1] - [V(0) - B] = \frac{\int_0^\infty [p - w(\tau)]N(\tau)d\tau}{R} - B_1 + B = 0 \quad (5.17)$$

where the zero follows from (5.11). This shows that if the separation rate remains the same under the new contract as it was under the old contract, then the value of a job to a new recruit will be unchanged and hence, from (5.12), profits will be the same. But, the mobility rule will not remain the same as wages are raised (assuming there are firms paying higher wages) and, as the mobility rule is determined by the worker to maximize the value of the job, it must be that the value of the job rises with the new contract leading to more recruits and higher profits.

Proof of Proposition 5.2

Continue to denote by $V(t)$ the value of the job at tenure t excluding any severance payments. Then, in the presence of severance payments, (5.1) is modified to

$$\delta_r V(t) = w(t) - \delta_u[V^u - V(t) - S(t)]$$

$$+\lambda \int_{V(t)+S(t)} [V - V(t) - S(t)]dF(V) + V'(t) - \delta_r S(t) \qquad (5.18)$$

where we have used the fact that workers will only now move to other firms when the offer V exceeds $[V(t) + S(t)]$. Profits (5.4) will now be

$$\Pi = \int_0^\infty [[p - w(t)]N(t) - S(t)N'(t)]dt + BN(0) \qquad (5.19)$$

which takes account of the fact that $-N'(t)$ workers of tenure t leave, each of which must make the payment $S(t)$ to the employer. Integrating the term $-S(t)N'(t)$ by parts, we can write (5.19) as

$$\begin{aligned}\Pi &= \int_0^\infty [p - w(t) + S'(t)]N(t)dt + [B + S(0)]N(0)\\ &= \int_0^\infty [p - \tilde{w}(t)]N(t)dt + \tilde{B}N(0) \qquad (5.20)\end{aligned}$$

where we have used the transversality condition that $\lim_{t\to\infty} S(t)N(t) = 0$. (5.20) shows that replacing the contract with severance payments by the one without leads to the same level of profits as long as employment is exactly the same. Showing this requires showing that initial recruits and separation rates will be the same under the two contracts. Define $V(t)$ to be the value of the job with the contract $\{\tilde{w}(t), \tilde{B}(t)\}$ so that

$$\delta_r \tilde{V}(t) = \tilde{w}(t) - \delta_u[\tilde{V}(t) - V^u] + \lambda \int_{\tilde{V}(t)} [V - \tilde{V}(t)]dF(V) + \tilde{V}'(t)$$

implies

$$\delta_r \tilde{V}(t) = w(t) - \delta_u[\tilde{V}(t) - V^u] + \lambda \int_{\tilde{V}(t)} [V - \tilde{V}(t)]dF(V) + \tilde{V}'(t) - S'(t) \qquad (5.21)$$

Comparing (5.21) and (5.18), one can see that we must have $\tilde{V}(t) = V(t) + S(t)$ which implies that separation rates must be the same and recruits must be the same. Hence, expected profits must be the same.

Proof of Proposition 5.3

We use a technique of proof similar to that used in Proposition 5.1 to show that replacing an arbitrary contract by one in which wages are initially at zero level and then, after some period, jump to p can only increase profits. Consider an initial contract $\{w(t)\}$. Consider replacing it by a contract which pays zero in the period $[0, t_0]$ and p thereafter where t_0 is chosen to satisfy

$$\int_{t_0}^\infty pN(t)dt = \int_0^\infty w(t)N(t)dt \qquad (5.22)$$

where $N(t)$ is employment on the initial contract. t_0 must exist and be strictly positive as long as the employer is making positive profits on the original contract. Using (5.22) we can write the difference between profits

on the new contract, Π_1, and on the old contract, Π, as being

$$\begin{aligned}\Pi_1 - \Pi &= \int_0^{t_0} pN_1(t)dt - \Pi \\ &= \int_0^{t_0} p[N_1(t) - N(t)]dt + \int_0^{\infty} pN(t)dt - \int_{t_0}^{\infty} pN(t)dt - \Pi \\ &= \int_0^{t_0} p[N_1(t) - N(t)]dt + \int_0^{\infty} [p - w(t)]N(t)dt - \Pi \\ &= \int_0^{t_0} p[N_1(t) - N(t)]dt \qquad (5.23)\end{aligned}$$

where $N_1(t)$ is employment on the new contract. So, if it can be shown that employment never falls with the new contract, then profits cannot fall either, and if there is a strict increase in employment, then profits must also rise. Whether employment rises or not depends on whether the value of the job rises or not.

We use the same technique as in Proposition 5.1 and start by assuming that the worker mobility decision remains the same under the new contract as it was under the old contract. Then, using (5.1), (5.13), (5.14), and (5.22), we have

$$V_1(t) - V(t) = \int_t^{\infty} [I(\tau, t_0)p - w(\tau)]N(\tau)d\tau \qquad (5.24)$$

where $I(\tau, t_0)$ is an indicator function taking the value 1 if $\tau \geq t_0$ and zero otherwise. For $t = 0$, we must have $V_1(t) = V(t)$. For $\geq t > 0$, we must have $V_1(t) \geq V(t)$. But, again the mobility rule will change with the new contract so as to increase $V_1(t)$ which means that employment can never decrease. Hence, profits can never decrease with the new contract and, if there is any effect on the value of the job, the new contract will actually increase it.

6

Earnings and the Life Cycle

SINCE at least the work of Mincer (1962, 1974) earnings functions have been an essential part of the toolbox of labor economists. These earnings functions are typically cross-section regressions of some measure of the wage or earnings on worker characteristics such as experience, job tenure, education and training, sex, race (even beauty and sexual orientation), and employer characteristics (for a survey, see Polachek and Siebert 1992). Estimating the returns to education, the extent of discrimination and diagnoses of the causes of rises in wage inequality are just some of the uses to which earnings functions have been put. The relationship between wages and gender is considered in chapter 7 and the relationship between wages and employer characteristics in chapter 8. The aim of this chapter is to understand the role of job search in the observed returns to experience and job tenure, or, put more broadly, the life-cycle profile of earnings.

This chapter differs from previous ones in that it is concerned with the behavior of workers rather than that of employers and treats the behavior of employers as given. From the perspective of workers, an oligopsonistic labor market presents itself as one in which there is wage dispersion, that is, there are good jobs and bad jobs and the worker wants to try to get him/herself into one of the good jobs, a process that might be described as job shopping. The analysis that follows then applies to any labor market in which there is wage dispersion although virtually all models of wage dispersion have, at their heart, a mechanism that gives employers some market power.

The cross-sectional relationship between earnings and experience for American and British men is presented in figure 6.1 (using data from the CPS and the LFS, respectively). In this chapter we focus exclusively on men because the observed cross-sectional profile for women is also heavily influenced by cohort effects resulting from the increase in the labor market participation of women. The basic facts about the profile are well known. Hourly earnings are a concave function of experience reaching a maximum at 25–30 years of experience for men with a modest decline towards the end of the working life. The experience profile is steeper in the United Kingdom than the United States with maximum gains of 120 log points for British men and 70 log points for the Americans. This is

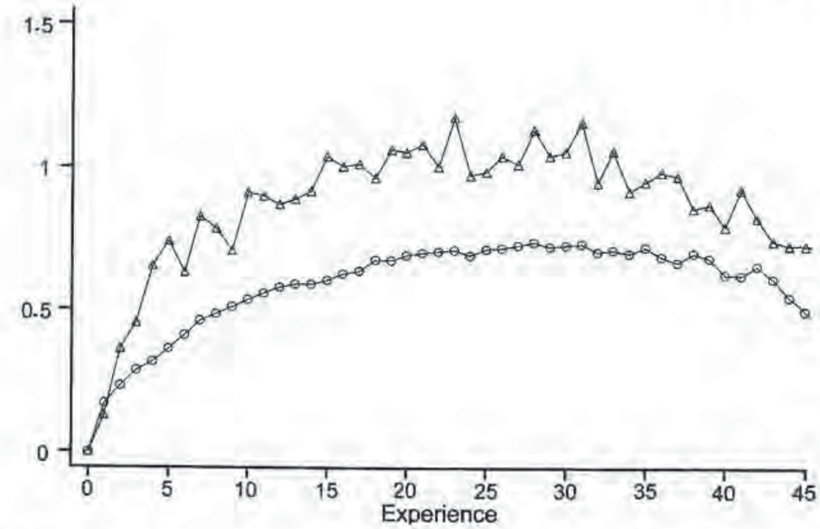

Figure 6.1 The returns to experience for American (○) and British (△) men.
Notes. The vertical axis is the gain in log hourly wages relative to workers with zero experience. Data sources are the CPS, January 1998 to December 1999, for the United States and the LFS, December 1992 to November 1999, for the United Kingdom. See Data Set Appendix for details of construction of experience. Data points are cell means. The US profile is smoother than the UK profile because of the larger size of the cells.

largely the result of faster wage growth in the first five years in the labor market in the United Kingdom than in the United States. In empirical work, the profile is often represented by a quadratic, but as Murphy and Welch (1990) emphasize, this is not really a good approximation (they suggest the use of a quartic).

The conventional way to understand the shape of the life-cycle profile of earnings is based on human capital theory and derives from the work of Mincer (1962) and Becker (1993). The profile observed in figure 6.1 is made up partly of true returns to experience and partly the returns to job tenure (average job tenure is, unsurprisingly, higher for more experienced workers). The returns to experience are interpreted as returns to general human capital (net of any current investment in this capital; see Mincer 1974; Ben-Porath 1967) while returns to job tenure are interpreted as a share in the returns to firm-specific human capital (although we criticized the foundation of this argument in the previous chapter). This chapter shows that the empirical evidence is not all supportive of the conventional interpretation of the observed returns to job tenure and experience (there is an empirical literature on the returns to tenure, discussed further below, that also reaches this conclusion; see Altonji

and Williams 1997). Restricting oneself to the human capital perspective misses out on important aspects of the way earnings evolve over the life-cycle.

This chapter argues for a more general way of thinking about the returns to experience and job tenure based on the returns to job mobility and the cost of job loss. It shows that traditional measures of the returns to tenure are weighted averages of the changes in returns to job mobility and costs of job loss. This way of thinking about the life-cycle profile of earnings is consistent with the way in which people perceive their lives; job changes and job loss are often perceived as "big" changes. Job-shopping models predict there will be cross-sectional correlations between wages, experience, and job tenure even if the evolution of wages within jobs is unrelated to these variables. Burdett (1978) pointed out that the more time one has spent searching for a job, the more likely it is that one has found a good one: hence average wages are likely to rise with experience. And, once a worker has found that dream job, they are less likely to leave it so their job tenure is likely to be long. As we shall see, the actual relationships predicted by the job-shopping model are more subtle than this discussion might suggest but we should not be surprised that there is a relationship at all.

The plan of this chapter is as follows. In the next two sections, we present two pieces of evidence that there is something wrong with the conventional human capital interpretation and specification of cross-sectional earnings functions. Section 6.1 examines the earnings losses of displaced workers, those individuals who have lost jobs through no fault of their own. It shows that earnings losses are related not just to previous tenure on the job but also to the level of experience, something we should not see if the returns to experience are the returns to general human capital. Also, the earnings of displaced workers are positively related to tenure on their previous job, something that is inconsistent with the view that the returns to job tenure are the returns to firm-specific human capital.

Section 6.2 shows that a considerable part of the returns to experience and tenure observed in a cross-section is the result of the fact that those who remain in a job are not randomly selected. The implication is that cross-sectional returns to tenure cannot be used as an estimate of the earnings growth that an individual can expect if they remain in their job.

Section 6.3 considers whether a job-shopping model can do better than the human capital model in explaining these empirical findings. We show that the job-shopping model can readily explain many of them but others are not entirely consistent with the pure search model. A more general framework for thinking about the life-cycle profile of earnings is necessary.

Section 6.4 proposes a new way of decomposing the life-cycle profile of earnings into earnings growth on-the-job, a cost of job loss, and a return to job mobility. The conventional measure of the return to job tenure is shown to be a function of the cost of job loss and the return to job mobility. We argue that this definition makes sense if all the returns to experience are returns to general human capital, and returns to tenure are returns to specific human capital, but we have already presented evidence that they are not. In this case, our decomposition is preferable.

In the final two sections, we present two applications of this approach. In section 6.5 we present some estimates of the returns to job mobility and, in section 6.6, we use the approach to try to understand the decline in earnings among older workers seen in figure 6.1.

6.1 The Earnings Losses of Displaced Workers

There are many studies (e.g., for surveys, see Jacobson et al. 1993a,b; Kletzer 1998) of the earnings of displaced workers, those workers who have lost their jobs through plant closures (which is taken to be involuntary on their part). As Jacobson et al. (1993a: 26) write "much academic research on displacement has examined whether human capital theory can account for the observed earnings reduction following dislocation." Human capital theory predicts that a worker experiencing job loss suffers some loss in his/her specific human capital but does not lose any general human capital. As specific human capital is presumed to be embodied in the observed returns to job tenure and general human capital in the returns to experience, the theory predicts that earnings losses should be associated with tenure on the previous job but not with experience. A large part of the literature on displaced workers has focused on showing that earnings losses are positively related to previous job tenure (e.g., Topel 1991; Farber 1997). But virtually all studies of displacement also find that, contrary to the standard interpretation of earnings functions, losses are also related to experience. This casts doubt on the human capital interpretation of the returns to experience. In addition, Kletzer (1989) and Neal (1995) find that earnings of displaced workers are related to tenure on the previous job, casting doubt on the hypothesis that all these are returns to firm-specific human capital.[1]

Let us look at some evidence on the magnitude of the earnings losses suffered by workers after displacement. In the United States we are fortu-

[1] Neal (1995) also finds that the impact of pre-displacement experience and tenure on the change in earnings are greater for those who change industries.

nate to have an occasional supplement to the CPS which is explicitly designed to focus on the fortunes of displaced workers: the Displaced Worker Survey (DWS). We use two of these surveys which were conducted in 1996 and 1998 although the results in Farber (1997) and other papers would indicate that the results presented also hold for the earlier surveys.

In this survey, workers are asked whether, in the previous three calendar years, they have lost or left a job "because a plant or company closed or moved, your position or shift was abolished, insufficient work or another similar reason," that is, the intention is to identify job loss that occurs for reasons beyond the control of the worker. Those workers who have experienced job loss are then asked about the job they lost, about how long it took them to find work again, and about their current job (if they are in employment). The structure of the survey and the fact that it is based on respondents' recall mean that it is difficult to use the survey to make calculations of the incidence of job loss (Evans and Leighton 1995) so that one cannot really use estimates of earnings loss from this survey to make inferences about the loss that would be suffered if the "average" worker lost their job. However, we have a more modest aim: to look at the way in which earnings losses are associated with job tenure and experience. Approximately 5% of the sample report being displaced in the previous three years. The average earnings loss is of the order of 10%. Properly measured, it should be larger than this as an average of 18 months had elapsed since job loss and earnings would have normally grown in that period, partly due to inflation and partly due to the normal life-cycle evolution of earnings.

Table 6.1 presents some earnings loss equations that are essentially the same as other papers in this area. We experimented with the inclusion of higher order terms in tenure and experience but they were never significant. One should note that the earnings on the current and displaced job are at two different points in time so that one would not expect the earnings to be the same even if there was no cost of job loss: for this reason we also include the time elapsed since job loss. Earnings losses are greater if the worker had more seniority on the previous job with each year of tenure leading to a loss of 1.2% in weekly earnings. It should be noted that this is much less than the return to tenure observed in the cross-section, the implication being that earnings on the new job are positively related to tenure on the previous job (Kletzer (1989) and Neal (1995) document this for earlier DWS) suggesting that not all returns to tenure are returns to firm-specific human capital.

It is also true that older workers suffer a greater earnings loss: a worker with 20 years of experience loses 14% more than a worker with no years of experience. Note also that the constant in these regressions can be interpreted as an estimate of the earnings loss for a worker for whom

TABLE 6.1
Estimates of Earnings Loss from Displacement: United States

	(1) All	(2) All	(3) Men	(4) Women
Previous tenure	−0.012	−0.014	−0.012	−0.014
	(0.0017)	(0.0025)	(0.0021)	(0.0030)
Previous experience	−0.007	−0.008	−0.008	−0.006
	(0.001)	(0.001)	(0.001)	(0.002)
Female	−0.041	−0.071		
	(0.021)	(0.031)		
White	−0.007	−0.070	−0.043	0.033
	(0.032)	(0.050)	(0.043)	(0.049)
<High school	0.015	−0.102	0.004	0.031
	(0.038)	(0.058)	(0.045)	(0.069)
High school graduate	−0.026	−0.067	−0.058	0.018
	(0.025)	(0.038)	(0.033)	(0.039)
College graduate	−0.017	−0.116	−0.042	0.027
	(0.028)	(0.040)	(0.035)	(0.046)
Dummy for 1996	0.025		0.025	0.024
	(0.010)		(0.013)	(0.016)
Years since displacement	0.183	0.184	0.172	0.187
	(0.073)	(0.112)	(0.093)	(0.116)
Years since displacement	−0.046	−0.050	−0.041	−0.050
squared	(0.022)	(0.033)	(0.028)	(0.035)
Previously union	−0.121	−0.150	−0.104	−0.163
	(0.032)	(0.048)	(0.038)	(0.059)
Change of industry		−0.122		
		(0.037)		
Current tenure		0.017		
		(0.005)		
Constant	−0.031	0.228	0.027	−0.146
	(0.494)	(0.098)	(0.080)	(0.097)
Number of observations	4293	1688	2478	1815
R^2	0.05	0.08	0.05	0.04

Notes.
1. The data used in these regressions come from the 1996 and 1998 displaced worker supplements to the CPS. The dependent variable is the log of the ratio of the post- to the pre-displacement hourly wage.
2. Standard errors are in parentheses.

all the regressors are zero, that is, the earnings loss for a worker in the reference category (male, black, some college in 1994) with zero years of experience and tenure who gets a job immediately after displacement from a non-union job. The constant is small and insignificantly different from zero suggesting that such a worker does not suffer large earnings losses after displacement: it is the older, more senior, previously union workers who do. As the estimated impact of many individual characteristics on the earnings loss is small, this conclusion can be taken to apply to workers of all genders, race, and education. The second column also includes controls for whether the worker changed industry and their job tenure on the current job, questions that were only asked in 1996. This has essentially no effect on the coefficients on previous job tenure and experience. The remaining columns of table 6.1 present separate estimates for men and women. The earnings losses of more experienced women do seem slightly less than those of more experienced men but this needs to be put into the context of the fact that the earnings profile is flatter for women than for men so that there is less to lose.

We have looked so far only at earnings losses after displacement. But it is important to remember that there are other ways in which older workers seem to suffer more after job loss: they seem to struggle to find any acceptable job at all. Estimates of probit models (not reported) for whether an individual has worked at all since job loss indicate that, controlling for other factors, a worker with 10 years of prior experience is 7% less likely to have found work afterwards. To the extent that this means that older workers are more selective from among the options available to them, the estimates of earnings losses in table 6.1 are probably an understatement of the extent of the problem facing older workers. One plausible interpretation of their plight is that their reservation wage is high relative to their earnings opportunities after displacement because it is related in part to their previous earnings (through savings or wealth effects).

This evidence suggests very strongly that, controlling for other relevant factors, more experienced workers suffer greater earnings losses on displacement. While this observation may not come as a surprise, it is not consistent with the interpretation of the returns to experience as the returns to general human capital. And, the fact that not all returns to tenure are lost also casts doubt on the interpretation of the returns to tenure as the returns to firm-specific human capital.

6.2 Sample Selection in the Cross-Sectional Earnings Profile

This section demonstrates that a considerable part of the returns to experience and job tenure observed in cross-section earnings functions

is the result of the sample selection bias that arises from the fact that (as predicted by the job-shopping model) workers with higher wages are more likely to stay in their jobs. Cross-section earnings functions are usually written in such a way as to give us a relationship between average log wages, experience (a), and job tenure (t): let us denote this relationship by $w(a, t)$. However, for reasons that will become apparent, it is more convenient here to focus on average earnings as a function of the experience level when the job started (which we denote by a_0) and job tenure. Let us denote this earnings profile by $\omega(a_0, t)$. Obviously we must have $\omega(a_0, t) = w(a_0 + t, t)$ so that there is a simple one-to-one relationship between the two profiles.

The advantage of conditioning on starting experience is that those workers with a level of job tenure, t, must, in some sense, be the subset of those workers with the same level of starting experience and job tenure zero who have remained in their jobs for t periods. Of course, in a cross-section they are not literally the same workers but, if the economy was in a steady state, they would be equivalent.

If all those who started jobs at experience a_0 remained in their jobs for t years, there would be no bias in the measured increase in earnings, $[\omega(a_0, t) - \omega(a_0, 0)]$. Note that this includes both the returns to job tenure and experience. The source of the potential bias is that not all workers will survive to have job tenure of t: some of them will lose their jobs and others will move to other jobs. If these moves are random then, again, there will be no bias in the tenure profile but if the selection of workers who survive to have tenure of t is not random, then $[\omega(a_0, t) - \omega(a_0, 0)]$ will give a biased estimate of the true increase in earnings. To give an example, suppose there is no wage growth on the job but that workers with high initial wages are more likely to remain in the same job. We would then observe that $\omega(a_0, t)$ is higher than $\omega(a_0, t-1)$ but this is entirely the result of the selection bias.

The measured return to an extra year on the job can be written as

$$\omega(a_0, t) - \omega(a_0, t-1) = [\omega(a_0, t) - \omega^s(a_0, t-1)] + [\omega^s(a_0, t-1) - \omega(a_0, t-1)] \quad (6.1)$$

where $\omega^s(a_0, t-1)$ is the average log wage at starting experience a_0 and job tenure $(t-1)$ for those workers who remain in the same job for one more year (who we call the stayers). The first term in square brackets is the wage growth for those who remain in the job. The second term in square brackets is the selection bias that arises because those who stay in the same job may not be randomly selected; we call this the stayer bias. Only the first term in square brackets can be used as an estimate of the wage growth expected for those who remain in their jobs.

With cross-section data we cannot estimate the extent of the stayer bias, but, with only rudimentary panel data one can, as the different terms in (6.1) are all readily measured. The left-hand side can be estimated using an observed cross-section and the stayer bias by comparing the lagged wages of the stayers and the non-stayers. (6.1) refers only to the one-year returns, but we can write the returns to t years of job tenure as

$$\begin{aligned}\omega(a_0,t)-\omega(a_0,0) &= \sum_{t'=0}^{t-1}[\omega(a_0,t-t')-\omega(a_0,t-t'-1)]\\ &= \sum_{t'=0}^{t-1}\Big[[\omega(a_0,t-t')-\omega^s(a_0,t-t'-1)]\\ &\quad +[\omega^s(a_0,t-t'-1)-\omega(a_0,t-t'-1)]\Big] \qquad (6.2)\end{aligned}$$

so that the cross-sectional return can be written as the cumulated wage growth for stayers and the cumulated stayer bias.

To measure the extent of the stayer bias we use the Panel Study of Income Dynamics (PSID), 1985–97, the National Longitudinal Survey of Youth (NLSY), 1980–94 for the United States and the British Household Panel Study (BHPS), 1992–98, for the United Kingdom. For the study of the evolution of wages over the life-cycle, the PSID has the big disadvantage that wage information is only collected for heads of household and their spouses. As relatively few individuals are in either category when they first enter the labor market, the sample size is small in the first years after labor market entry when earnings typically grow the fastest. The NLSY does not have this problem as it is explicitly focused on young workers: the disadvantage is that we have no observations on older workers. The BHPS does not have either of these problems as wage information is recorded for all individuals in sample households.

We could present the decomposition in (6.2) using the actual cell means observed in our data sets. But, the relatively small sample sizes mean that some experience–tenure cells are extremely small and the components of (6.2) show a lot of sampling variation. Given this, we model the various components by estimating a quartic in starting experience and job tenure with a full set of interaction terms and using the fitted values as the components in (6.2). Because of the linearity of the decomposition in (6.2) the resulting estimates are internally consistent. To estimate the cross-section profile (the left-hand side of (6.2)), we simply estimate an earnings function using current wages. To estimate the stayer bias (the final term on the right-hand side of (6.2)), we estimate an earnings func-

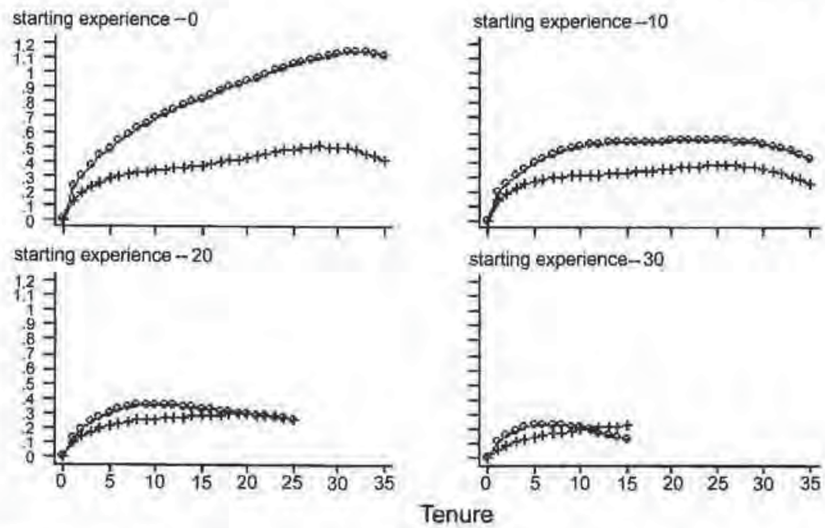

Figure 6.2 The stayer bias and the cross-sectional earnings profile: US men (PSID). O, cross-sectional earnings profile; +, cumulative stayer bias.

Notes. This figure represents the decomposition of the return to tenure of (6.2). The earnings profile is the estimated observed return to tenure in the cross-section (the left-hand side of (6.2)), and the stayer bias is the cumulative bias because those who stay are not randomly selected (the second term on the right-hand side of (6.2)).

tion for the lagged wage for all workers and for stayers, the difference between them being an estimate of the stayer bias. The wage growth for stayers must then be difference between the two.[2]

The results for the PSID are reported in figure 6.2 for years of starting experience equal to 0, 10, 20, and 30 years. The profiles all normalize the earnings of workers with zero tenure and a given level of starting experience to be zero. The measured cross-sectional returns to remaining in a job are large—for example, 50 log points for a worker with 10 years starting experience and 10 years of job tenure. But most of this is estimated to be the result of the stayer bias. The greater the starting experience, the larger the fraction of the cross-section returns that seem to be the result of the stayer bias. One might be concerned that these results are

[2] There is a good reason for having the wage growth for stayers as the residual. Nominal and real wages typically increase for reasons independent of the individual. In estimating wage growth for stayers, we would not want to include this aggregate wage growth. In some studies, an external measure of average wages is used to de-trend wages in the sample. This is satisfactory if one can be sure that the sample is representative of the population but quite small departures can make quite large differences given that average wage growth is not large. De-trending using within-sample information is better but still problematic and it is probably best avoided if possible (as here).

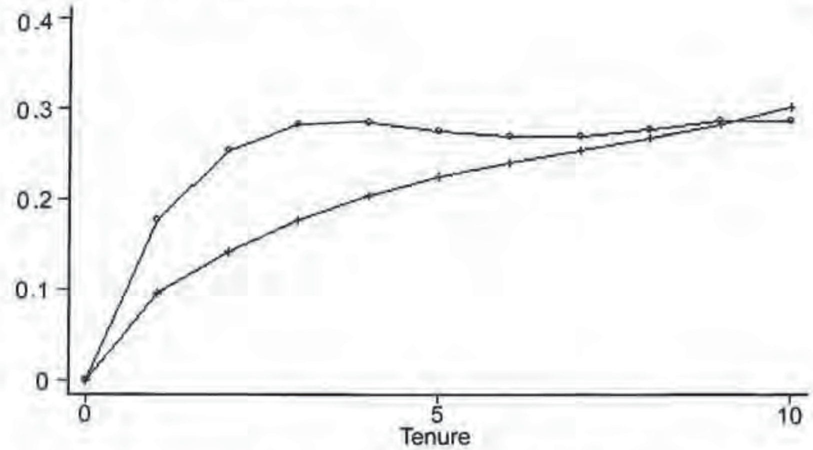

Figure 6.3 The stayer bias and the cross-sectional earnings profile: US men (NLSY). O, cross-sectional earnings profile; +, cumulative stayer bias.
Notes. This is drawn for a starting experience of zero. As for figure 6.2.

driven by the small number of observations in the PSID of individuals with low levels of experience. But the results for men with zero years of starting experience from the NLSY reported in figure 6.3 suggest a similar conclusion. Finally, figure 6.4 shows similar results for British men from

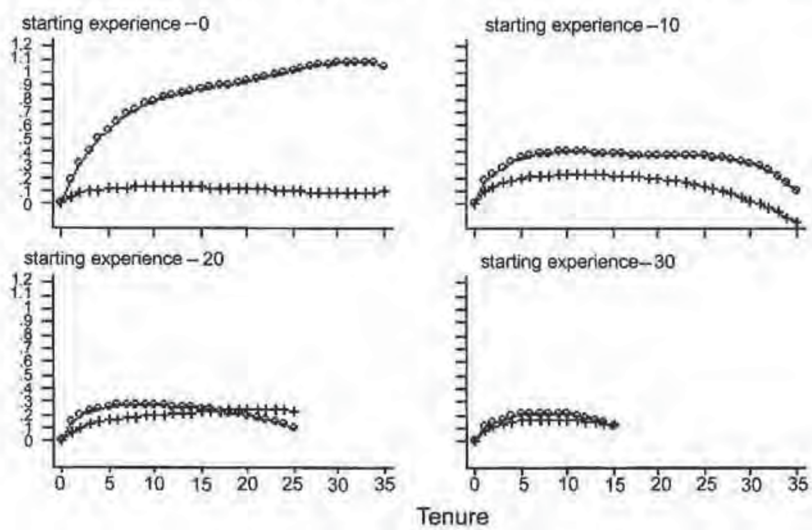

Figure 6.4 The stayer bias and the cross-sectional earnings profile: UK men (BHPS). O, cross-sectional earnings profile; +, cumulative stayer bias.
Notes. As for figure 6.2.

the BHPS although the importance of the stayer bias is much less among those starting jobs when young.

This section has shown that cross-sectional earnings functions suffer from serious bias caused by the fact that those in higher wage jobs are more likely to remain in them. Although we have shown the existence of serious problems with the traditional interpretation of earnings functions, it is not obvious that a job-shopping model can do better: this is the subject of the next section.

6.3 The Cross-Sectional Returns to Experience and Tenure in a Job-Shopping Model

This section discusses the predictions of a job-shopping model for the observed cross-sectional relationship between wages, experience, and job tenure. Because it is easy to get bogged down in the technicalities, the discussion in the main body of the chapter sticks to the intuition and all the technical material is confined to an appendix.

Individuals enter the labor market and try to work themselves into the better jobs through the process of job search. Every so often, an opportunity in the form of a job offer will arrive and the worker will take it if it is better than the current job. This process of systematically choosing better jobs means that, as workers age, we would expect to see average wages rising. However, job destruction sometimes interrupts this process and forces the worker to start again. A lifetime career then comes to resemble a game of "snakes and ladders" in which job offers represent ladders enabling workers to advance faster and job losses represent "snakes" which cause setbacks.

Now, consider a more formal model of this process. Denote job tenure by t and experience by a. In the language used in the discussion of earnings functions, this is really potential rather than actual experience but, for simplicity, we simply refer to it as experience.[3] Initially, assume that there are no "true" returns to experience and job tenure so that the only way in which wages can increase is by changing jobs: we call this the pure search model (for other discussions of this, see Burdett 1978; Manning 1998, 2000). In this case we can represent the wage-offer distribution by $F(w)$ as we have done in earlier chapters. In a later section, we consider the case where there are true returns to experience and job tenure. For simplicity, we assume that job offers arrive at a rate λ and job loss occurs at a rate δ as in the basic dynamic oligopsony model of section 2.4.

[3] One could do the analysis that follows conditioning on actual experience rather than potential although the analysis is messier. As most studies of earnings functions do not have data on actual experience, the case analyzed seems the more important in practical terms.

Three main cases are considered:

- the distribution of wages conditional on experience alone when there are no true returns to experience or job tenure (the pure search model);
- the distribution of wages conditional on experience and job tenure when there are no true returns to experience or job tenure (the pure search model);
- the distribution of wages conditional on experience and job tenure when there are true returns to experience or job tenure.

6.3.1 The Relationship Between Wages and Experience in a Pure Search Model

As we saw in figure 6.1, earnings profiles for men are typically a concave function of experience, first increasing and then decreasing. It is natural to ask whether the pure search model can explain these stylized facts. The following result summarizes Proposition 6.2, presented in more detail in the appendix. It turns out that the fraction of employment that is made up of recruits from non-employment has an important role to play in explaining the shape of the wage–experience profile.

Result 6.1. *The pure job-shopping model has the following predictions:*

1. *For low enough levels of experience, expected wages will be an increasing, concave function of experience.*
2. *A sufficient condition for expected wages to be increasing in experience is that the share of recruits from non-employment in total employment is non-increasing in experience for all lower levels of experience.*
3. *A necessary condition for expected wages to be decreasing in experience is that the share of recruits from non-employment in total employment is increasing in experience for some lower levels of experience.*
4. *A sufficient condition for expected wages to be a concave function of experience is that the share of recruits from non-employment in total employment is constant at all lower levels of experience.*

Proof. See Proof of Proposition 6.2 in Appendix 6A.

These results establish that the pure search model can go some way towards explaining the stylized facts of the earnings profile. Part 1 shows

that it predicts that the earnings function must initially be an increasing, concave function of experience as we observe. The intuition for why the search model predicts this profile is that older workers tend to have worked themselves into the better jobs through the process of job search and there are diminishing returns to the search for better and better jobs. One way to understand this prediction is to think about what the profile looks like in heaven where nobody ever loses their job and everybody lives forever. Then, workers can only ever move up the job ladder so that average wages must be increasing, but all workers eventually end up in the best-paying job at which point there is no scope for increasing wages further.

But the wording of the result also makes it clear that the pure search model does not always predict that earnings will be an increasing function of experience. In particular it predicts that if, in some phase of the life-cycle, the fraction of employment recruited from non-employment rises, then we might expect to see declining average earnings. The reason is that these workers recruited from non-employment are likely to be entering employment at relatively low wages so a high fraction of these workers will tend to drag down average wages. It is interesting to test this prediction by seeing whether there is evidence for an increase in the fraction of employment previously non-employed among older workers where average earnings decline.

Figure 6.5a plots on the same graph (but using a different scale) the relationship between wages and experience and the fraction of those currently in employment who were previously in employment (previously being a month ago) for the United States. We plot the fraction previously in employment simply because this makes clearer the positive relationship between this profile and the earnings profile. This figure is not very clear as it is dominated by the rapid change in both series in the first five years after labor market entry. Figure 6.5b presents the same graph but ignoring the youngest workers (note that the scales are different from those presented in figure 6.5a). The close relationship between the two series should be clear. The earnings profile begins to turn down as the fraction of the employed previously employed falls. Figure 6.6a,b does the same exercise for British data. Again, we observe a similar picture.[4] This is consistent with a search view of the labor market as outlined in Result 6.1.

The pure search model can also explain the earnings losses suffered by more experienced displaced workers. Earnings rise with experience because workers have had longer to find a good job: involuntary displacement results

[4] Note that the fraction of those currently in employment who were previously non-employed is higher in the United Kingdom than the United States because "previously" refers to three months ago rather than a month, a result of the fact that the UK LFS is quarterly and the US CPS is monthly.

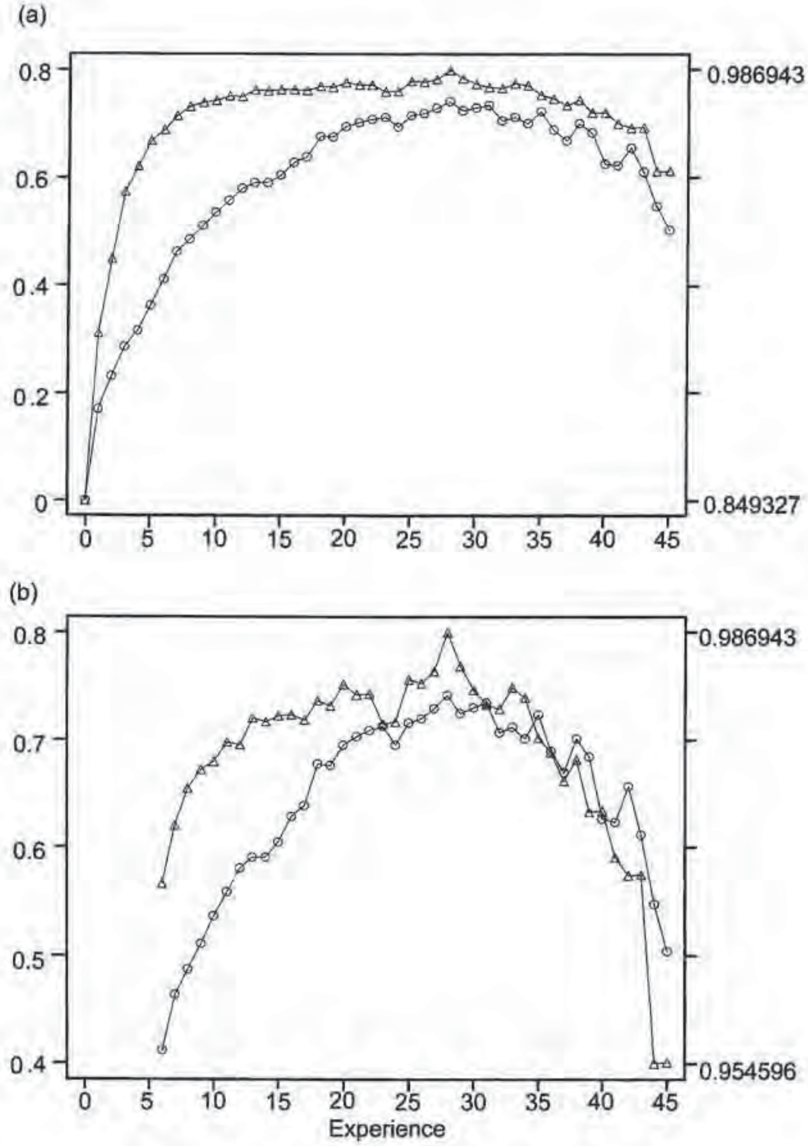

Figure 6.5 The connection between the earnings profile and the previously employed profile in the United States (CPS data). (a) All workers; (b) excluding youngest workers. O, mean log hourly wage (left scale); Δ, proportion previously employed (right scale).

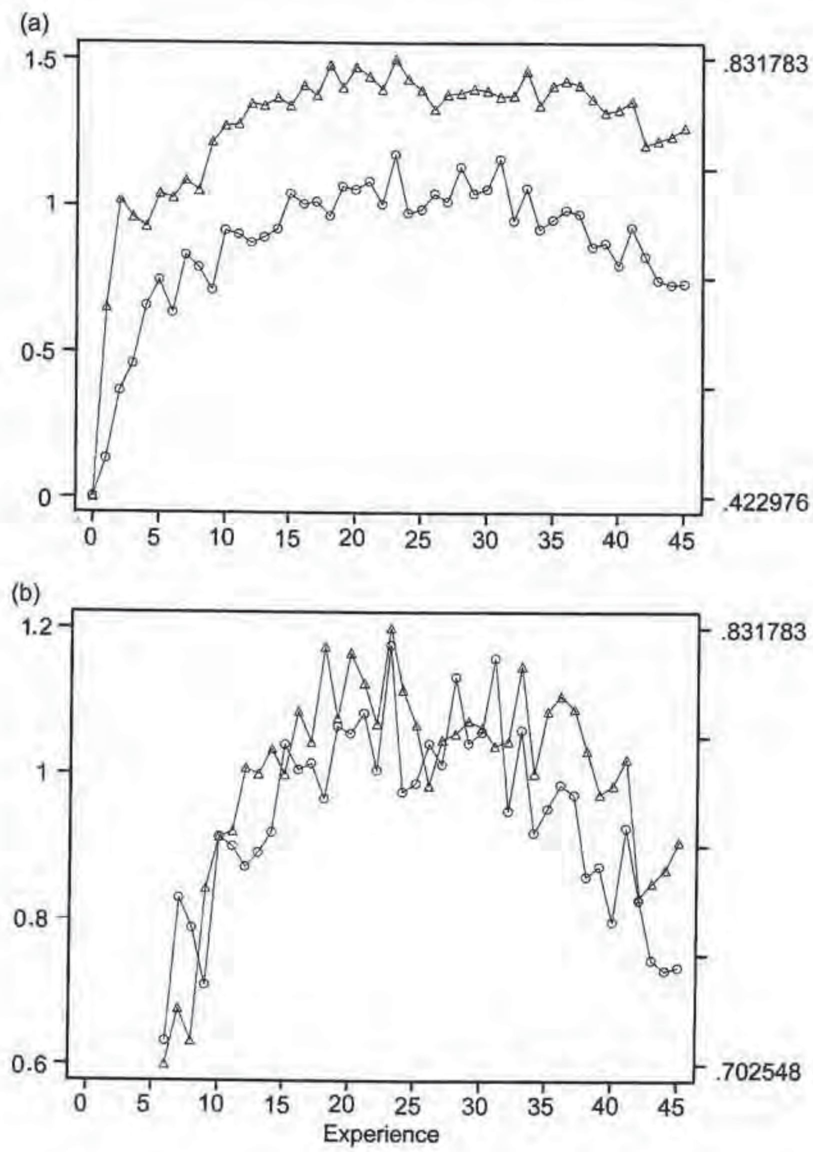

Figure 6.6 The connection between the earnings profile and the previously employed profile in the United Kingdom (LFS data). (a) All workers; (b) excluding youngest workers. ○, mean log hourly wage (left scale); △, proportion previously employed (right scale).

in the loss of that job, forcing the worker to start looking for a good job again. One can think of earnings as being related to "search capital," knowledge about the location of good jobs. Unlike general human capital, some search capital is immediately destroyed on displacement. In the simple model presented here, all search capital is destroyed and the earnings of displaced workers are independent of experience. However, one might think that more experienced workers do retain some established contacts and knowledge about the labor market so that, in reality, not all search capital is destroyed immediately after displacement.

6.3.2 *The Relationship Between Wages, Experience and Tenure in a Pure Search Model*

Now introduce an extra conditioning variable, job tenure, into the analysis. First, consider the relationship between tenure and expected wages. The basic conclusion to remember is that the pure search model has an ambiguous prediction about the returns to job tenure. There are a number of ways of representing this ambiguity but one is summarized in the following result.

Result 6.2. *Expected wages are increasing (decreasing) in job tenure as*

$$R^{u}(a-t)(R^{u}(a-t)+\lambda)+\frac{\partial R^{u}(a-t)}{\partial a}>(<)0 \tag{6.3}$$

where $R^{u}(a-t)$ *is the fraction of recruits from non-employment in total employment at experience* $(a-t)$.

Proof. See Proof of Proposition 6.3 in Appendix 6A.

Expected wages are correlated with job tenure in the pure search model for two reasons. First, a job tenure of t tells us that no better job offer than the current one has arrived in length of time t. As better job offers are less likely to arrive when the current job pays a high wage, this tends to induce a positive correlation between tenure and wages. It is this mechanism that probably lies behind the commonly expressed view that job shopping induces a positive correlation between job tenure and the wage. But, there is also another effect at work. If we also condition on the current experience of the worker, then current job tenure tells us the experience level at which the worker started the current job (which is $a-t$). If we compare two workers with the same experience but with one year's difference in tenure, the more senior worker must have started the job one year earlier in their career than the junior worker. If expected wages for workers starting jobs are increasing with experience, this effect will tend to make us

predict that the senior worker had a lower starting wage than the junior worker. Result 6.2 simply gives us a condition when we can sign the net effect although the link between (6.3) and the intuition is not very clear.

One can also use (6.3) to ask the question "when does an earnings function give an unbiased estimate of the returns to tenure?" In this case the "true" returns to tenure are zero and, by inspection of (6.3), one can see that this condition is satisfied in the situation where there are never any recruits from non-employment.[5] In this case, experience is a sufficient statistic for expected wages and the two effects of job tenure exactly cancel out as emphasized in Topel (1991). But, in general, there is no reason to believe that the cross-sectional returns to tenure will be unbiased. One conclusion to be drawn is that it is the length of time in continuous employment that is a sufficient statistic for wages in this simple model: this insight is used below.

Now, consider the relationship between expected wages and experience when job tenure is also a conditioning variable. One might expect that one could reproduce something like Result 6.1 and readily provide sufficient conditions for the average wage to be an increasing function of experience. But, the introduction of tenure as an extra conditioning variable means that it is possible that, conditional on tenure, the expected wage can be declining in experience over some region even when the conditions of Result 6.1(2) are satisfied.

To understand the reason for this, suppose we observe two sets of workers with different levels of experience but with identical job tenure. We can divide each of these sets of workers into two groups:

1. those that arrived in the present job directly from another job;
2. those that arrived in the present job from non-employment.

The distribution of wages among these two groups will be different but, if we condition only on experience and tenure, the observed wage distribution will be a mixture of the two. As the wage distribution shifts up with experience (if we assume the sufficient conditions of Result 6.1(2)) the wage distribution must be higher for an older worker from group 1 than for a younger worker. In contrast the wage distribution among those from group 2 must be independent of experience as displacement is assumed to result in the loss of all search capital. In addition, the wage distribution among group 2 must be less than that among group 1. If the proportion of workers from the two groups did not vary with experience, then the fact that the wage distribution of the first group increases with experience and the wage distribution of the second group is independent of experience would mean that the distribution of the mixture of the two must be increasing with experience. But, the proportion does change and it is

[5] This is a sufficient, not necessary, condition.

Figure 6.7 Variation in the fraction of recruits from non-employment. ○, US (PSID); △, UK (BHPS).
Notes. The left-hand scale is the fraction of recruits (i.e., those workers with tenure zero) who were in employment a year ago.

possible that this can go some way towards explaining why average earnings decline for older workers. Figure 6.7 shows, using data from the PSID for the United States and the BHPS for the United Kingdom that the fraction of mature male recruits (i.e., those with short job tenure) who were previously in employment falls with experience. When we observe an old male worker with low job tenure, it is quite unlikely that they arrived in this new job because they got a better job offer and relatively likely that they arrived in the new job after an intervening period of non-employment.

So far we have analyzed the consequences of experience and tenure for the wage distribution in a search model in which the true returns to experience and job tenure are zero. But, it is also instructive to consider the case in which the true returns are not zero.

6.3.3 The Distribution of Wages With "True" Returns to Experience and Tenure

In this case the model is modified in the following way. First, we assume that the wage offer distribution changes with experience so that the wage offer distribution at experience a can be written as $F(w - \beta_a(a))$ where

$\beta_a(a)$ is a measure of the true returns to experience.[6] We normalize $\beta_a(a)$ so that $\beta_a(0) = 0$. We also assume that log wages in a job rise at a rate β_t with tenure. The assumption that the returns to tenure are linear is restrictive but it helps to make the analysis tractable. Its advantage is that expected wage growth on two jobs is independent of the current wage and tenure so the choice between them can be made solely on the basis of which job offers the highest wage.[7]

Note that we are saying nothing about the sources of returns to experience and tenure. It may be that the traditional interpretation is correct and these are the returns to general and specific human capital. But it may be that they are the results of the wage policies of discriminating monopsonists as suggested in the previous chapter.

Understanding all the effects at work in this case is complicated and not a task to be undertaken lightly. Perhaps the only simple result is the obvious one that the addition of a true return to experience simply raises the predicted return to experience by that amount. This is unsurprising: as earnings on the job and outside offers evolve mechanically according to the function $\beta_a(a)$; there is no extra source of bias induced here.

More interesting is the return to tenure. Result 6.2 provided the condition (6.3) for the cross-sectional return to tenure to be unbiased when the true return to tenure is zero. In the appendix, it is shown that this condition being satisfied does not guarantee that the observed return to tenure is equal to the true return when there are non-zero returns to tenure (i.e., $\beta_t \neq 0$). The sign of the bias depends on the functional form of the wage offer distribution so that it is hard to form any a priori view about it. Given that the theoretical effects at work are complex, perhaps a helpful way of seeing the different effects at work is through simulation: table 6.2 presents four cases. We consider different values of the parameters β_t representing the "true" return to tenure and δ_u the rate of job loss. For convenience we assume that on-the-job and off-the-job search are equally effective, that the "true" return to experience is zero, and that the log wage offer distribution is logistic with mean zero (this will be the average log wage of a worker with zero experience). Case I is the case where there is no "true" return to tenure and no risk of job loss: Result 6.2 then applies and the return to tenure is zero in this case (read down a column). However, the apparent return to experience is sizeable (read across a row). Case II now introduces a risk of job loss (sufficient to make the

[6] We assume that the wage offer distribution improves mechanically with experience and does not depend on actual time in employment. This is consistent with empirical practice in which we generally do not observe "true" labor market experience. An obvious alteration to the model is to assume that the wage offer distribution only improves with time actually spent in employment: analytically this is much messier than what follows.

[7] See Topel (1986) and Mortensen (1988) for a discussion of the mobility rule for workers in the case where returns to tenure are not linear.

TABLE 6.2
Simulated Returns to Experience and Tenure

Years of job tenure	*Years of experience*			
	10	*20*	*30*	*40*
Case I. No true return to tenure ($\beta_t = 0$), no risk of job loss ($\delta_u = 0$)				
0	0.19	0.27	0.32	0.34
10		0.27	0.32	0.34
20			0.32	0.34
30				0.34
Case II. No true return to tenure ($\beta_t = 0$), risk of job loss ($\delta_u = 0.1$)				
0	0.09	0.09	0.09	0.09
10		0.24	0.24	0.24
20			0.3	0.3
30				0.33
Case III. True return to tenure ($\beta_t = 0.01$), no risk of job loss ($\delta_u = 0$)				
0	0.22	0.35	0.45	0.5
10		0.38	0.47	0.55
20			0.49	0.58
30				0.6
Case IV. True return to tenure ($\beta_t = 0.01$), risk of job loss ($\delta_u = 0.1$)				
0	0.09	0.08	0.08	0.08
10		0.32	0.32	0.32
20			0.45	0.45
30				0.56

Notes.
1. The other assumptions in these simulation are that $\lambda = 0.5$, that all workers are initially in unemployment, and that the wage offer distribution is logistic with mean zero and standard deviation of 0.2.

steady-state non-employment rate 28%). Now, the return to experience, conditional on job tenure, is essentially zero but there is a large return to job tenure. The reason is that job tenure is now a much better measure of how long a worker has been in continuous employment than experience. Case III goes back to assuming there is no job destruction but assumes the true return to tenure is 1% per year. Compared to Case I, there is now a return to job tenure in the cross-section but it is much less than the "true" return. This is an important point: search models do not always predict a higher return to tenure in the cross-section than the true tenure effect. Topel (1991: 151) was the first to make this point in the form of an example (in which there was never any unemployment so the job destruc-

tion rate was zero) but, unfortunately, he then went a bit too far and claimed that "the basic theory of search and matching implies that ... a comparison of wages for workers with different job tenures will *understate* the returns to seniority." This is not as widely true as he claimed. Another point about Case III worth noting is that the cross-sectional returns to experience are much larger than in Case I: part of the returns to tenure are "captured" in the return to experience. It is easy to understand why. Suppose we compare two workers both with zero tenure but who differ in one year of experience. The older worker could have remained in the same job in which case earnings would have grown by β_t but chose to move because a better job came along. Their earnings in this case must have grown by more than β_t. The returns to experience will then appear to be β_t while the return to tenure will actually appear to be negative. Finally, the fourth case has a true return to tenure and some risk of job loss. The returns to experience are very small while the returns to tenure are large and well above the true returns.

One use of this type of simulation is to suggest an interpretation of the negative returns to seniority often found in academic labor markets (see, e.g., Ransom 1993; Moore et al. 1998; Bratsberg et al. 2002). In academic labor markets the risk of job loss for senior faculty is plausibly thought of as low: as a result the first and third cases might be thought to be the most relevant. As we have seen, the search model there predicts that the cross-section will be an underestimate of the true return to tenure. It would be interesting to see whether other professions with similar low risks of job loss also display negative returns to job tenure.

This section and the previous one have emphasized that the search framework provides reasons for why cross-sectional returns to experience and tenure are likely to be biased. One important message is that a search approach suggests that time since last non-employed is important in determining earnings. If workers rarely leave employment then experience rather than tenure is better correlated with this variable so we would expect to see large returns to experience. But, for workers who often have spells of non-employment, job tenure might be better correlated. So, the search model tends to predict that workers with weak labor market attachment have higher returns to tenure: a human capital approach would not predict this as the incentives to invest in specific human capital are weak for these groups. We discuss this further in chapter 7 in the context of the cross-sectional returns to tenure being higher for women than men.

The two previous sections have argued that there is something wrong with the conventional human capital interpretation of earnings functions. This section has considered whether a search model can explain stylized facts about the distribution of wages conditional on experience and job tenure. We have shown that the search model has the potential to explain:

- the concavity in the experience profile of earnings;
- the decline in earnings for older workers;
- the earnings losses of more experienced displaced workers discussed in section 6.1;
- the biases in the cross-sectional returns to tenure discussed in section 6.2.

However, there are some aspects of the profile that the search model cannot explain. For example, section 6.2 pointed out that the earnings of displaced workers are positively related to previous job tenure. In addition, this section has been entirely theoretical, the theoretical predictions are often ambiguous, and theory provides little guidance for how one should try to understand the life-cycle of earnings in practice. The next section develops a more practical approach.

6.4 Empirical Approaches to the Estimation of the Life-Cycle Profile in Earnings

There is a large literature on estimating the "true" returns to tenure and experience (see Topel 1986, 1991; Abraham and Farber 1987; Altonji and Shakotko 1987; Marshall and Zarkin 1987; Altonji and Williams 1997, 1998; Teulings and Hartog 1998). The basic approach can be understood very simply.

Suppose we observe an individual in employment at experience 0 with log earnings that we normalize to zero as represented in figure 6.8a. The following year, we observe them with experience 1 but they might be in the same job or a new job (for simplicity ignore the possibility that they might not be in employment). The log earnings associated with the two possibilities are also represented in figure 6.8a. The return to experience is then estimated as the earnings growth on new jobs, that is, the distance OA and the return to tenure is estimated as the extra increase in earnings for those who remain in the same job, that is, the return to a year of tenure is estimated as AB. Of course, estimating these two returns is not as straightforward in practice as it appears in figure 6.8a as one only observes one of the two outcomes at experience 1 in figure 6.8 and one has to worry about sample selection, etc. Correcting for these problems is the main pre-occupation of the papers on estimating the returns to experience and tenure cited above.

Figure 6.8a is derived from Becker's (1993) view of the reasons for the returns to experience and job tenure in which returns to experience reflect increases in general human capital and returns to job tenure reflect increases in specific human capital. In this case, the difference (in terms of human capital) between those who remain in their job and those who

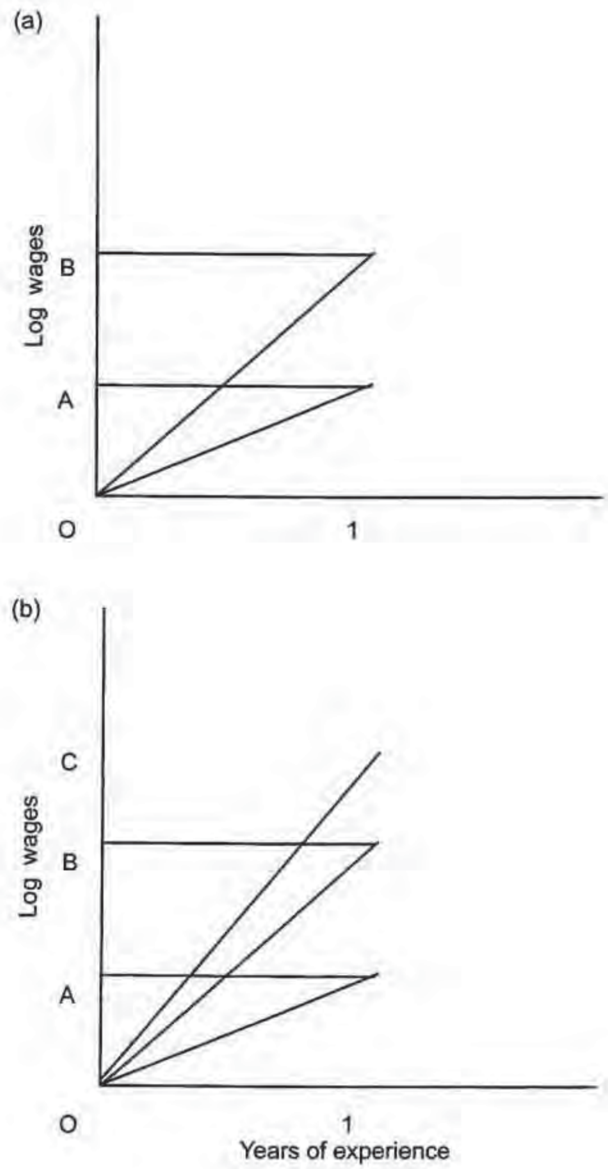

Figure 6.8 (a) The returns to experience and job tenure in the human capital model. (b) The evolution of earnings in the search model.

do not is the accumulation of one year of specific human capital (in addition to any accumulation of general human capital) so it makes sense to define the distance AB as the returns to job tenure.

However, figure 6.8a makes less sense if one uses the job-shopping model as the basis for analysis of the evolution of earnings. It then makes sense to divide those who change jobs into two groups: those who have got a better job offer and have moved voluntarily, and those who lost their initial job involuntarily and were forced to take a new job. We would expect those who move jobs voluntarily to have earnings above those who remain in their old job, and those who move involuntarily to have lower earnings. A possible picture is represented in figure 6.8b where those who change jobs voluntarily have higher earnings (OC) than those who remain in their jobs (OB), and those who lose their jobs have lower earnings (OA).

Figure 6.8b could be reduced to figure 6.8a by grouping together all those who are in new jobs: OA in figure 6.8a would then be a weighted average of OA and OC in figure 6.8b, the weight being the proportion of those in new jobs who moved involuntarily. Having reduced figure 6.8b to figure 6.8a one could then define the return to experience and job tenure in the usual way. But, there is a loss of information in doing so and it is not clear that it makes much sense. For example, an increase in the rate of job destruction would reduce the measured returns to experience and increase the measured returns to tenure, not because anything in figure 6.8b changes but because the fraction of displaced workers in those starting new jobs has changed.

If one started with figure 6.8b in mind, one could summarize it as:

1. the growth in earnings on-the-job, OB;
2. the cost of job loss, AB;
3. the return to job mobility, BC.

It makes sense to think in terms of these three components. In fact, if one asks a worker to describe how their earnings evolved over their life-cycle, it is quite likely that they would give an answer in terms of these concepts.

This result that the measured return to job tenure will be a weighted average of the returns to job mobility and the costs of job loss extends to more than the two-period model represented in figures 6.8a and 6.8b.

For example, to keep things simple assume that log wages grow on the job at the same rate, g, for everyone, that everyone faces a probability δ of losing one's job with an associated wage gain of Δ^l (that is probably negative), that everyone has a probability λ of getting a new job with an associated wage gain of Δ^m. The return to an extra year of job tenure can be written as

$$w(a,t) - w(a,t-1) = [gt + w(a-t,0)] - [g(t-1) + w(a-t-1,0)]$$

$$\times w(a-t,0) - w(a-t+1,0) + g \qquad (6.4)$$

After some algebra one can derive that, for $a > 1$, the return to an extra year of job tenure is given by

$$w(a,t) - w(a,t-1) = \lambda\Delta^{m} + \delta\Delta^{l} \qquad (6.5)$$

so that the costs of job loss and the returns to job mobility continue to be important.

There is already literature on estimating the costs of job loss (or displacement) (for surveys, see Jacobson et al. 1993a; Kletzer 1998) and the returns to job mobility or job shopping (Topel and Ward 1992). This literature co-exists with that on the returns to experience and job tenure but, as the above discussion should make clear, these concepts cannot all be independent of each other.

In the next two sections, we investigate a couple of the issues that might usefully be addressed using the framework proposed here: the return to job mobility and the declining earnings of older workers.

6.5 Estimating the Return to Job Mobility

In this section we present some estimates of the returns to job mobility using the PSID and NLSY for the United States and the BHPS for the United Kingdom. There are some papers which attempt to estimate the returns to job mobility. Topel and Ward (1992), in a sample of US young men, found an average wage gain at transition of about 10%, roughly double what the earlier studies of Mincer (1986) and Bartel and Borjas (1982) had found.

We are interested in the returns to voluntary job mobility but we do not observe directly whether job moves are voluntary or not. In all of our data sets, we define a job move as voluntary if there was no intervening period of non-employment and involuntary if there was. This classification is likely to make mistakes. Workers may have some advance warning of impending job loss but manage to find another job before they actually lose the current one (see, e.g., Jones and Kuhn 1995): we will classify these incorrectly as voluntary job moves. And some workers may arrange for a period of non-employment (a "break") between finishing an old job and starting a new one: these will be classified incorrectly as involuntary job moves. Assuming that wage growth is a motive for some job moves and that there is a cost of job loss, this inability to identify the reasons for job moves means we are

likely to understate both the returns to job mobility and the costs of job loss. But, given the data available, the classification adopted here seems the best one.

Let us consider some estimates of the returns to job mobility in our data sets. Tables 6.3–6.5 show results for the PSID, NLSY, and BHPS, respectively. The first column of each table shows the coefficient on a dummy variable for whether an individual was a mover or not. The sample is workers who have been in continuous employment so all those for whom mover is zero are those who have remained in their jobs. For all our data sets, movers have significantly higher average wage growth than stayers: 5.1% in the PSID, 1.7% in the NLSY, and 4.3% in the BHPS. In these regressions, the constant is the average real wage growth for stayers: this is 2.6% in the PSID, 1.3% in the NLSY, and 2.1% in the BHPS so job mobility is estimated to double or even triple wage growth. These estimates do not control for any other characteristics. The second column introduces lagged experience and job tenure. Job tenure is lagged because it must, by definition, be zero for those who moved jobs over the year: it represents the accumulated tenure on the previous period's job. Experience is included in lagged form just for comparability. After some experimentation with higher-order polynomials, we settled on a specification with a quadratic in lagged experience and a linear term in lagged tenure. In all the data sets, this reduces the estimated gains from job mobility: to 3.9% in the PSID, 1.0% in the NLSY, and 2.4% for the BHPS. The reason is simple: those with low tenure and experience are more likely to move and these groups tend to have higher wage growth. Given that wage growth varies with tenure and experience, we might think that the gains from job mobility also vary so the third column includes interactions of mover with lagged job tenure and experience. The current returns from job mobility are declining in both experience and job tenure in all three data sets and the tenure effects are larger than the experience effects.

Why might the returns to mobility be declining in tenure? There are a number of possible explanations. The job-shopping model would suggest that, as job tenure increases, it is more likely that the worker is in one of the better jobs in which case the opportunities for wage gains from job mobility are reduced. Secondly, it may be the result of rational decision-making. If wage growth declines with job tenure (and there is some evidence of this), one of the returns to changing job is that the worker can expect faster wage growth on the new job than they would have had in the old job. A more senior worker might even be prepared to accept an instantaneous cut in wages in the expectation that future wage growth would be higher on the new

TABLE 6.3
The Returns to Job Mobility in the United States: PSID

	(1) *Movers + stayers*	*(2)* *Movers + stayers*	*(3)* *Movers + stayers*	*(4)* *Movers + stayers*	*(5)* *Movers + stayers*	*(6)* *Stayers*	*(7)* *Stayers*
Mover	0.051	0.039	0.073	0.072	0.080		
	(0.004)	(0.004)	(0.008)	(0.008)	(0.010)		
Mover × tenure (−1)/10			−0.068	−0.067	−0.071		
			(0.010)	(0.013)	(0.012)		
Mover × experience (−1)/10			−0.012	−0.024	−0.092		
			(0.005)	(0.005)	(0.006)		
Tenure (−1)/10		−0.0141	−0.0128	−0.0130	−0.0121	−0.0124	0.0027
		(0.0017)	(0.0018)	(0.0018)	(0.0049)	(0.0017)	(0.0033)
Experience (−1)/10		−0.0254	−0.0223	−0.0227		−0.0213	−0.0066
		(0.0043)	(0.0044)	(0.0045)		(0.0044)	(0.006)
Experience (−1)/10 squared		0.0039	0.0034	0.0038		0.0033	−0.0003
		(0.0010)	(0.0009)	(0.0010)		(0.0010)	(0.0013)
Constant	0.026	0.066	0.062				
	(0.001)	(0.004)	(0.004)				
Sample selection							0.140
							(0.030)
Personal characteristics	No	No	No	Yes	No	Yes	Yes
Fixed effects	No	No	No	No	Yes ($p = 1.00$)	No	No
Time effects	No	No	No	Yes	Yes	Yes	Yes
Number of observations	53053	53053	53053	53052	53053	48926	37555
R^2	0.003	0.007	0.008	0.013	0.010	0.007	–
Average gain to job move	0.051	0.039	0.040	0.039	0.050	0.039	0.313

Notes.
1. The dependent variable is the change in the log hourly wage.
2. The sample consists of workers who are in continuous employment from one year to the next.
3. Standard errors in parentheses.

TABLE 6.4
The Returns to Job Mobility in the United States: NLSY

	(1) *Movers+ stayers*	(2) *Movers+ stayers*	(3) *Movers+ stayers*	(4) *Movers+ stayers*	(5) *Movers+ stayers*	(6) *Stayers*	(7) *Stayers*	(8) Stayers
Mover	0.017	0.010	0.028	0.029	0.046			
	(0.006)	(0.006)	(0.010)	(0.010)	(0.014)			
Mover × tenure (−1)/10			−0.056	−0.053	−0.033			
			(0.043)	(0.043)	(0.052)			
Mover × experience (−1)/10			−0.030	−0.033	−0.046			
			(0.018)	(0.018)	(0.024)			
Tenure (−1)/10		−0.031	−0.029	−0.034	−0.050	−0.036	0.107	0.071
		(0.014)	(0.016)	(0.016)	(0.025)	(0.014)	(0.152)	(0.039)
Experience (−1)/10		−0.116	−0.103	−0.054		−0.020	−0.096	−0.031
		(0.027)	(0.028)	(0.032)		(0.033)	(0.130)	(0.037)
Experience (−1)/10 squared		0.067	0.064	0.033		0.004	0.074	0.017
		(0.022)	(0.023)	(0.025)		(0.026)	(0.144)	(0.029)
Constant	0.013	0.052	0.047					
	(0.003)	(0.007)	(0.007)					
Sample selection							0.101	0.123
							(0.129)	(0.041)
Personal characteristics	No	No	No	Yes	No	Yes	Yes	Yes
Fixed effects	No	No	No	No	Yes ($p = 1.00$)	No	No	No
Time effects	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	16030	16030	16030	15971	16030	11551	2336	13507
R^2	0.001	0.004	0.004	0.005	0.005	0.006	–	–
Average gain to job move	0.017	0.010	0.011	0.011	0.023	0.011	0.174	0.206

Notes.
1. As for table 6.3.

TABLE 6.5
The Returns to Job Mobility in the United Kingdom: BHPS

	(1) *Movers + stayers*	*(2)* *Movers + stayers*	*(3)* *Movers + stayers*	*(4)* *Movers + stayers*	*(5)* *Movers + stayers*	*(6)* *Stayers*	*(7)* *Stayers*	*(8)* Stayers
Mover	0.043	0.024	0.086	0.087	0.095			
	(0.007)	(0.007)	(0.011)	(0.016)	(0.016)			
Mover × tenure (−1)/10			−0.131	−0.131	−0.116			
			(0.017)	(0.017)	(0.021)			
Mover × experience (−1)/10			−0.019	−0.019	−0.025			
			(0.006)	(0.007)	(0.009)			
Tenure (−1)/10		−0.014	−0.011	−0.012	−0.027	−0.0111	−0.0029	0.0127
		(0.003)	(0.003)	(0.003)	(0.012)	(0.0031)	(0.0050)	(0.0065)
Experience (−1)/10		−0.055	−0.046	−0.045		−0.048	−0.042	−0.011
		(0.006)	(0.006)	(0.007)		(0.006)	(0.009)	(0.010)
Experience (−1)/10 squared		0.0096	0.0079	0.0080		0.0085	0.0077	0.0022
		(0.0015)	(0.0015)	(0.0015)		(0.0015)	(0.0018)	(0.0021)
Constant	0.021	0.088	0.077					
	(0.002)	(0.006)	(0.006)					
Sample selection							0.068	0.250
							(0.043)	(0.060)
Personal characteristics	No	No	No	Yes	No	Yes	Yes	Yes
Fixed effects	No	No	No	No	Yes ($p = 1.00$)	No	No	No
Time effects	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of observations	17800	17800	17800	17699	17800	16032	14871	12801
R^2	0.004	0.013	0.018	0.020	0.006	0.010	–	–
Average gain to job move	0.043	0.024	0.027	0.028	0.027	0.027	0.150	0.489

Notes.
1. As for table 6.3.

job.[8] Thirdly, it may be the case that, for more senior workers, a higher proportion of job moves that we class as voluntary are, in fact, involuntary.

The fourth column introduces personal characteristics (sex, race, education, and region) and time dummies into the wage growth equations. While these variables are jointly significant, they have essentially no effect on the estimated returns to job mobility. While personal characteristics are very important in explaining differences in the level of wages, they are much less important in explaining the growth of wages among those who remain in employment. The fifth column goes further in considering whether unobserved heterogeneity can bias the returns to job mobility by introducing person-specific fixed effects. It has been argued that those with less stable work histories also tend to have lower wage growth on the job. There is some evidence for this in all of our data sets as the inclusion of fixed effects does raise the estimated return to job mobility. However, the effects are not large and, in all the data sets, one can accept the hypothesis that the person-specific fixed effects are jointly zero (the p values are reported).[9] The results here are rather different from those of Baker (1997) and Meghir and Pistaferri (2001) who emphasize the importance of growth-rate heterogeneity. Of course, this should not be taken to mean that unobserved heterogeneity is unimportant in explaining the *level* (rather than the growth) of wages and the coefficient on tenure in a cross-sectional wage equation is typically halved by the inclusion of fixed effects.

So far, all our equations have been estimated for movers and stayers together. But, a substantial part of the literature on the returns to tenure (e.g., Topel 1991) simply estimates a wage growth equation for stayers and then uses this to predict the wage growth that movers would have obtained if they had not changed jobs. The sixth column estimates a wage growth equation for stayers, then uses these estimates to predict what wage growth for the movers would have been if they had stayed and the average gain to a job move is then estimated as the gap between the actual

[8] Recognition of this point means we also have to be more careful in estimating the returns to job mobility. Suppose that a worker leaves a job with tenure t_0. Then, assuming he does not leave the new job and would not have left the old one, the gap in earnings after τ years will be given by $\Delta w_\tau = \beta_0 + [\beta_1 - \beta_2(\tau/10)](t_0/10)$ where β_0 is the coefficient on the mover dummy (0.073 in column 3 of table 6.3, β_1 is the coefficient on the mover dummy interacted with lagged tenure (−0.068 in column 3 of table 6.3), and β_2 is the coefficient on lagged tenure (−0.013 in column 3 of table 6.3). To give an example, suppose a worker leaves a job with accumulated tenure of 10 years. On starting the new job, the pay gain is estimated to be 2.5% but, five years later, the predicted pay gap from the NLSY between the new and old job is 3.7%.

[9] Note that one can no longer identify the experience effect once one has time dummies and fixed effects.

wage growth of movers and this predicted wage growth. These estimates are very similar to those obtained in earlier specifications for the simple reason that the estimated coefficients in the stayers equation are not very different from those we had obtained earlier. However, one concern with these estimates is that stayers and movers are not likely to be randomly selected so that sample selection issues arise in the interpretation of all the estimates we have discussed so far.

A simple model makes clear the potential source of the bias. Assume that, within jobs (i.e., for stayers), wage growth for individual i at date τ is given by

$$\Delta w_{i\tau}^{s} = \beta^{s}(a_{i\tau-1}, t_{i\tau-1}) + \beta_{x}^{s} x_{i\tau} + \varepsilon_{i\tau}^{s} \tag{6.6}$$

and that the wage growth if they moved would be

$$\Delta w_{i\tau}^{m} = \beta^{m}(a_{i\tau-1}, t_{i\tau-1}) + \beta_{x}^{m} x_{i\tau} + \varepsilon_{i\tau}^{m} \tag{6.7}$$

We also need to specify a mobility rule which tells us whether workers decide to stay or go. It is plausible to think that such a decision is based partly on the difference in wages between this job and the new job but other factors, for example, non-monetary aspects of jobs, are also likely to be relevant. So, let us define a latent index $S_{i\tau}^{*}$ according to

$$S_{i\tau}^{*} = \gamma_0[\Delta w_{i\tau}^{s} - \Delta w_{i\tau}^{m}] + \gamma_1(a_{i\tau-1}, t_{i\tau-1}) + \gamma_2 z_{i\tau} + \xi_{i\tau}^{s} \tag{6.8}$$

and assume that the individual stays in the job when $S_{i\tau}^{*} > 0$. This mobility rule could be derived from optimizing behavior and a number of papers do this (e.g., Topel 1986; Mortensen 1988) by assuming that workers maximize a value function that depends solely on monetary rewards. However, we do not want to put this much structure on the model.

The potential source of selection bias should be apparent. If wages are important in influencing job mobility decisions, then we would expect those individuals with a larger than expected innovation to the stayers' wage growth will be more likely to stay. There is a traditional way to deal with this potential problem (Heckman 1976; Lee 1978). This involves substituting (6.6) and (6.7) into (6.8) to yield the reduced-form stayers' equation

$$\begin{aligned} S_{i\tau}^{*} = {} & \gamma_0[\beta^{s}(a_{i\tau-1}, t_{i\tau-1}) - \beta^{m}(a_{i\tau-1}, t_{i\tau-1})] + \gamma_1(a_{i\tau-1}, t_{i\tau-1}) \\ & + \gamma_0(\beta_{x}^{s} - \beta_{x}^{m}) x_{i\tau} + \gamma_2 z_{i\tau} + \gamma_0[\varepsilon_{i\tau}^{s} - \varepsilon_{i\tau}^{m}] + \xi_{i\tau}^{s} \end{aligned} \tag{6.9}$$

and then typically assuming joint normality of the errors. Let us write (6.9) in the form

$$S_{i\tau}^{*} = \gamma z_{i\tau}^{s} + \xi_{i\tau} \tag{6.10}$$

where z^s is all the regressors from (6.9) and the variance of ξ is normalized to one. Then, as is well known, the expected wage growth for stayers, conditional on staying in the same job is given by

$$E(\Delta w_{i\tau}^s \mid S_{i\tau}^* > 0) = \beta^s(a_{i\tau-1}, t_{i\tau-1}) + \beta_x^s x_{i\tau} + \rho\sigma^s \frac{\phi(-\gamma' z_{i\tau}^s)}{1 - \Phi(-\gamma' z_{i\tau}^s)} \quad (6.11)$$

where σ^s is the standard deviation of ε^s and ρ is the correlation coefficient between ε^s and ξ. The last term is the sample selection correction term. If we are interested in estimating the returns to job mobility, we want to compare the wage growth of movers with the wage growth they would have achieved if they had stayed in their jobs. As we know they chose to move, this is given by

$$E(\Delta w_{i\tau}^s \mid S_{i\tau}^* < 0) = \beta^s(a_{i\tau-1}, t_{i\tau-1}) + \beta_x^s x_{i\tau} - \rho\sigma^s \frac{\phi(-\gamma' z_{i\tau}^s)}{\Phi(-\gamma' z_{i\tau}^s)} \quad (6.12)$$

So, our strategy is to use the standard Heckman approach to sample selection to get estimates of (6.11) and then use these estimates to estimate (6.12).

For the sample selection correction to work well, one needs variables that affect the stay–go decision but do not affect wage growth (the "z" variables in (6.9)).[10]

The BHPS and, in some years, the NLSY ask some questions about job satisfaction. For example, the BHPS asks questions about satisfaction with promotion possibilities, total pay, relations with the boss, job security, use of initiative, work itself, and hours worked while the NLSY asks questions about satisfaction with promotion possibilities, total pay, job security, ability to use and learn skills, relations with co-workers and supervisors, and the physical aspects of the job. In addition, the BHPS asks questions about whether domestic responsibilities have limited job search. We might expect some of these variables to be correlated with wage growth (e.g., satisfaction with pay, promotion possibilities, even relations with the boss) but there are some questions that we might expect to have no relationship to wage growth but do affect the appeal of the job. Our first strategy for investigating the extent of sample selection bias is to use these variables as the "z" variables. As plausibly exogenous "z" variables we used satisfaction with use of initiative, with work itself, and with hours worked for the BHPS, and satisfaction with co-workers for the NLSY. In addition, we used a question that asks whether domestic commitments have prevented job change for the BHPS. We first need to check that they are correlated with mobility. Table 6.6 presents some job

[10] Though an additional problem is that, if there is heterogeneity in the returns to job mobility, the use of an instrument will only give the "treatment effect on the treated."

TABLE 6.6
Job Mobility Equations

	(1) BHPS	*(2) NLSY*	*(3) BHPS*	*(4) NLSY*	*(5) PSID*
Tenure (−1)/10	0.579	3.354	0.536	2.774	0.656
	(0.038)	(0.293)	(0.041)	(0.088)	(0.025)
Experience (−1)/10	0.257	−1.014	0.056	−0.454	0.204
	(0.050)	(0.496)	(0.061)	(0.140)	(0.043)
Experience (−1)/10 squared	−0.032	1.092	0.019	0.301	−0.031
	(0.013)	(0.572)	(0.015)	(0.117)	(0.010)
Satisfaction with use of initiative	0.025				
	(0.012)				
Satisfaction with work itself	0.056				
	(0.013)				
Satisfaction with hours worked	0.047				
	(0.010)				
Satisfaction with co-workers		0.117			
		(0.054)			
Domestic constraints prevent job change	−0.206				
	(0.078)				
Log wage (−2)			0.230	0.186	0.214
			(0.042)	(0.029)	(0.025)
Personal characteristics	Yes	Yes	Yes	Yes	Yes
Time effects	Yes	Yes	Yes	Yes	Yes
Number of observations	14871	2336	12801	13509	38384
Proportion stayers	0.9	0.727	0.906	0.635	0.927

Notes.

1. The sample consists of those workers who have been in continuous employment for the past year. The dependent variable takes the value one if the individual is a stayer and zero otherwise. The estimated model is a probit. Standard errors in parentheses.
2. The sample size is much reduced for the NLSY as the detailed job satisfaction questions were only asked in 1979–82 and 1988.
3. For the NLSY, the job satisfaction variable ranges from 1 (not true at all that co-workers are friendly) to 4 (very true). For the BHPS, the job satisfaction variables range from 1 (not satisfied at all) to 7 (completely satisfied).

mobility equations. The first two columns show that, for both the BHPS and the NLSY, these variables are correlated with job mobility.

The 7th column of tables 6.4 and 6.5 then present estimates of wage growth equations for stayers with a sample selection correction term based on these mobility equations. In both cases, the coefficient on the sample selection correction term is significantly different from zero, and doing the sample selection correction makes a big difference to the estimated gain to job mobility which rises to 17 log points for the NLSY and 15 log points for the BHPS.

An alternative approach to sample selection correction is to use the lagged wage as the "z" variable in the movers equation. This is reasonable if the wage is correlated with job mobility but not with wage growth on-the-job, that is, wage growth on-the-job is a random walk with drift. Inclusion of the lagged wage in the stayers' wage growth equations (we used the specification of column 6 in tables 6.3–6.5) would not seem to support this assumption: the coefficient on the lagged wage being −0.182 (0.003) in the PSID, −0.232 (0.006) in the NLSY, and −0.193 (0.006) in the BHPS (standard errors in parentheses). But these coefficients are undoubtedly biased away from zero by the presence of measurement error in the wage. If we instrument the lagged wage using the second lag of the wage, the coefficients change to −0.029 (0.004) in the PSID, −0.038 (0.011) in the NLSY, and −0.034 (0.007) in the BHPS (standard errors in parentheses) so that the random-walk assumption does not seem such a bad assumption. So, we also experiment with the second lag of the wage as a "z" variable in the mobility equation. Columns (3)–(5) of table 6.6 indicate that the second lag is very significant in predicting job mobility.

Column (7) of table 6.3 and column (8) of tables 6.4 and 6.5 then estimate a wage equation for stayers that includes the sample selection correction term from these mobility equations: the coefficients on these terms are always significantly different from zero. The consequences for the estimated returns to job mobility are dramatic, rising to 31 log points in the PSID, 21 log points in the NLSY, and 49 log points in the BHPS. These are too large to be taken seriously as, given that the actual wage growth for movers is much more modest, they imply very negative wage growth for these workers if they had stayed with their firms. It may be that the functional form assumption is inappropriate[11] (particularly for

[11] Topel (1986) took a similar approach to sample selection correction, but Topel (1991) abandoned it suggesting that it was not reliable. However, his proposed "solution" was to assume there was no sample selection bias in a wage growth equation for stayers. He claimed some evidence in support of this but it was not very persuasive as wage innovations have a strong permanent component and the level of wages is strongly correlated with job mobility.

the PSID and BHPS where the share of movers in the sample is very small so that we are in the tail of the distribution), that wage growth within jobs is not a random walk, or that there is heterogeneity in the returns to job mobility and low-wage workers have potentially large gains.

From this discussion, one should conclude that there is some evidence that sample selection bias is important in estimating the returns to job mobility. But, because it is hard to find good instruments for job mobility, it is hard to get a precise estimate of the extent of the bias. But, including the job satisfaction variables does indicate that the return to job mobility may be closer to 10% than the raw figures reported in tables 6.3–6.5.

6.6 The Life-Cycle Profile of Earnings for Older Men

In this section we aim to put together the framework described in section 6.4 to understand one interesting feature of the earnings profiles seen in figure 6.1, the decline in earnings for older workers. The following proposition shows how one can decompose the change in earnings into the wage growth for continuing workers, the returns to job mobility, the costs of job loss, and one other term which we will call the leaver bias.

Proposition 6. *If the average log wage for workers with experience a is $w(a)$ then*

$$w(a+1) - w(a) = \overline{\Delta}^{s}(a) + \theta^{m}(a+1)\overline{\Delta}^{m}(a) + \theta^{e}(a+1)\overline{\Delta}^{l}(a+1) + \frac{n(a+1)-n(a)}{n(a+1)}\overline{b}^{l}(a) \qquad (6.13)$$

where $\overline{\Delta}^{s}(a)$ is the average log wage growth for those who stay in their jobs, $\theta^{m}(a+1)$ is the fraction of those currently in employment who have changed jobs in the past year, $\overline{\Delta}^{m}(a)$ is the average gain from job mobility, $\theta^{e}(a+1)$ is the fraction of those currently in employment who have come via non-employment in the past year, $\overline{\Delta}^{l}(a)$ is the average cost of job loss, $n(a)$ is the employment rate for experience a, $\overline{b}^{l}(a)$ is the "loser bias", the extent to which those who lose jobs are paid less than the average.

Proof. See Appendix 6B (where precise definitions of the different terms are also provided).

The decomposition in (6.13) may look intimidating but it says that the return to an extra year of experience can be thought of as being made up of several components:

- wage growth on-the-job, $\overline{\Delta}^s(a)$;
- the return to job mobility, $\overline{\Delta}^m(a)$, multiplied by the fraction of workers who change jobs, θ^m;
- the cost of job loss, $\overline{\Delta}^l(a)$, multiplied by the fraction of employed workers who have entered from non-employment, θ^e;
- the bias which emerges because those losing jobs are not selected at random.

Most of these terms are self-explanatory: wages can grow over the life-cycle because they grow for those in continuous employment, many workers have changed jobs and/or there is a return to job mobility or because few have lost jobs. However, the loser bias term perhaps needs a little bit more explanation. Suppose there was no on-the-job wage growth, no return to job mobility or cost of job loss, but that low-wage workers are more likely to leave employment and, once they have done so, never return to employment. Then, there will be a sample selection effect that will make wages appear to rise with experience. Trying to estimate the size of this effect is the subject of a recent paper by Blundell et al. (1999). But, it can only have an effect to the extent that the employment rate changes over time.

Let us now try to use this identity to understand the reasons why earnings decline for workers with experience in the range of 25–45 years. For this purpose, we use the PSID for the United States and the BHPS for the United Kingdom. (6.13) presents the one-year returns to an extra year of experience but we cumulate these up to present estimates of $w(a) - w(25)$. Inspection of figure 6.9a,b shows the decline in average earnings for older men that we see in the cross-section profile. However, those workers who manage to remain in their jobs actually see small gains in earnings (this is the cumulative wage growth for stayers) so that workers who manage to remain in their jobs do not experience wage declines on average at the end of their careers: this sounds plausible. Gains from job mobility are very small, partly because very few older workers change jobs and partly because the measured gains from job mobility are small for this group. The decline in earnings can be fully explained by the substantial incidence of job loss (something like 8–9% of those currently in employment have had a spell of non-employment in the last year) and the very substantial cost of job loss (measured at about 30 log points). It is the incidence and cost of job loss that accounts for the apparent decline in earnings of older men.

It should be emphasized that (6.13) is simply a decomposition that, because it is an identity, must hold true at all points in time. To move beyond the decomposition and ask questions like "if there was a higher rate of job mobility what would the profile look like?" one needs to make further assumptions, for example, that Δ^s and Δ^m are unaffected.

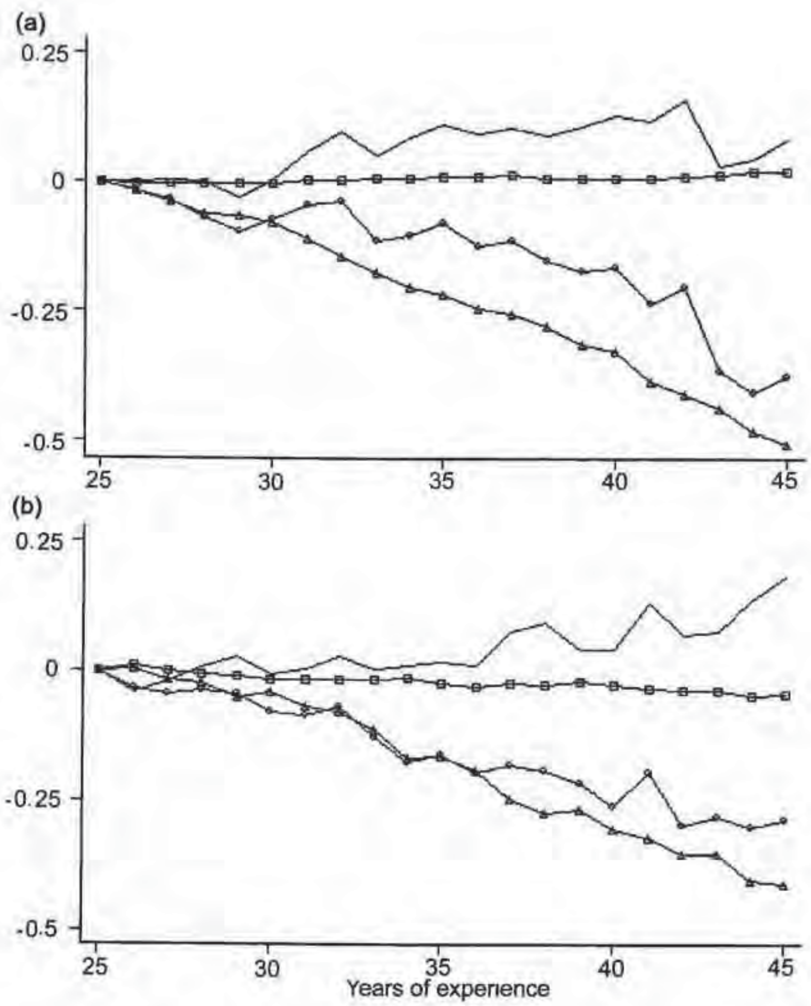

Figure 6.9 The decline in earnings for older men. (a) United States (PSID); (b) United Kingdom (BHPS). ○, change in average log wages; △, cumulative cost of job loss; □, cumulative job mobility gain; solid line, cumulative stayers wage growth.

Notes. This figure presents the decomposition of (6.13) cumulated over the last 20 years of labor market experience. The bias term of (6.13) is not presented but is very small.

6.7 Conclusions

This chapter has argued that there is something seriously wrong with the traditional interpretation of the returns to experience and job tenure in cross-sectional earnings functions. The evidence on displaced workers in section 6.1 makes it clear that not all of the returns to experience are the returns to general human capital and the evidence on the stayer bias in section 6.2 makes it clear that a large part of the apparent returns to job tenure is the result of the fact that those in high wage jobs are more likely to remain in them.

Section 6.3 then argued that a job-shopping model is a much better way to think of the way in which earnings evolve over the life-cycle. In a labor market with wage dispersion and frictions, one can think of an individual's progress through their working life as a game of snakes and ladders. If they remain in the same job their earnings will grow (although typically at a decreasing rate). But, from time to time, a "ladder" appears in the form of an opportunity to change jobs and get the wage gains that typically accrue from this mobility (although we have argued that it is difficult to estimate these gains precisely). But, there are also the snakes of job loss which typically result in sizeable earnings losses. In the early years of one's life, finding a "ladder" is more important than avoiding a "snake" (as was shown in section 6.5) but the latter is more important for older workers (as was shown in section 6.6). This representation of a working life coincides with the way that life is experienced.

Section 6.4 proposed a better way of thinking about the shape of earnings profiles: a decomposition into earnings growth within jobs, the returns to job mobility, and the costs of job loss. It was shown how the return to tenure as conventionally measured is a weighted average of the return to job mobility and the cost of job loss so compounds two very distinct concepts in a way that is unhelpful.

Our approach to understanding the profile of earnings requires panel data. But, in many situations, we have only cross-section data and this then raises the practical question of what labor economists should do when they want to estimate an earnings function. This chapter has implications not just for how labor economists should interpret earnings functions but also for how they should specify them.

There is an established "custom and practice" in the specification of earnings functions. It is not really clear whether this custom and practice is thought of as the "best specification" or the "best specification given the data available." There is perhaps a tendency to think in terms of the former when the latter would be more accurate. For example, given data

on wages, experience, and tenure, most labor economists would estimate a standard earnings function with these variables as regressors. Suppose that your particular data set also had information on whether the worker was previously employed when the individual started the current job. I suspect that this variable would be very significant partly because of unobserved individual heterogeneity but also because those who were previously employed are almost certainly changing jobs because they have found a "better" job. The traditional specification of earnings functions lumps together as equivalent those who have lost their previous job involuntarily and have re-entered the labor market and those who have voluntarily changed jobs. But we have seen (and this is hardly surprising) that there are substantial differences in the labor market fortunes of the two groups. Of course, there are some benefits to having a common specification of earnings functions across data sets but there are also dangers that this leads to a "lowest common denominator" outcome. It is likely that lots of variables summarizing the working life of an individual are important predictors of earnings (for an example of this, see Light and Ureta 1995): job tenure should be just thought of as one such measure, one that is commonly available but has no particular significance attached to it. Hence, an additional message of this chapter is that labor economists should be a bit more broad-minded in both the specification and interpretation of earnings functions.

This chapter has looked at the earnings profiles for men on the grounds that the observed female cross-sectional profiles are likely to be heavily contaminated by cohort effects. The next chapter aims to redress this balance and discuss how monopsony can help us explain the differences in earnings between men and women.

Appendix 6A. Theoretical Analysis of the Returns to Experience and Job Tenure in a Search Model

The Pure Search Model

Initially, assume that there are no "true" returns to experience and job tenure so that the wage offer distribution, $F(w)$, is independent of a or t. Denote by $G(w|a)$ the fraction of workers in employment of experience a who earn less than w. Knowledge of $G(w|a)$ is sufficient to work out statistics about the wage distribution, for example, the average wage $E(w|a)$ can be written as

$$E(w|a) = \int w dG(w|a) = \int [1 - G(w|a)]dw \qquad (6.14)$$

where the second equality follows from integration by parts. Using (6.14), we have

$$\frac{\partial E(w|a)}{\partial a} = -\int \frac{\partial G(w|a)}{\partial a} dw \tag{6.15}$$

so that $[\partial G(w|a)/\partial a] < 0$ for all w is a sufficient condition for average wages to be increasing in experience. The condition $[\partial G(w|a)/\partial a] < 0$ is simply the condition that increasing experience leads to a higher wage distribution in the sense of first-order stochastic dominance.

As earnings profiles are often concave in experience (see figure 6.1) we might also be interested in the second derivative of $E(w|a)$. Further differentiation of (6.15) leads to

$$\frac{\partial^2 E(w|a)}{\partial a^2} = -\int \frac{\partial^2 G(w|a)}{\partial a^2} dw \tag{6.16}$$

The following proposition provides an expression for $G(w|a)$.

Proposition 6.1. *$G(w|a)$ must satisfy the following differential equation*

$$\frac{\partial G(w|a)}{\partial a} = -\lambda[1 - F(w)]G(w|a) + [F(w) - G(w|a)]R^u(a) \tag{6.17}$$

where λ is the arrival rate of job offers and $R^u(a)$ is the flow of recruits from non-employment as a fraction of current employees.

Proof. Denote by $M(w|a)$ the number of individuals of experience a that are paid w or less. Obviously $N(a)G(w|a) = M(w|a)$ where $N(a)$ is employment of workers of experience a.

Workers leave the group of individuals paid less than w when they leave employment or they get a better job offer, that is, at a rate $[\delta + \lambda(1 - F(w))]$. Workers can only be recruited to the group of workers paid w or less from non-employment. Using the notation of Proposition 6.1 there are $N(a)R^u(a)$ recruits from non-employment of which a fraction $F(w)$ will be paid w or less. The change in $M(w|a)$ will then be the difference between inflows and outflows so will be given by

$$\frac{\partial M(w|a)}{\partial a} = -[\delta + \lambda(1 - F(w))]M(w|a) + F(w)N(a)R^u(a) \tag{6.18}$$

Now we must have $N(a)G(w|a) = M(w|a)$ so that

$$\frac{\partial M(w|a)}{\partial a} = N(a)\frac{\partial G(w|a)}{\partial a} + G(w|a)\frac{\partial N(a)}{\partial a} \tag{6.19}$$

and the change in employment must be given by

$$\frac{\partial N(a)}{\partial a} = -\delta N(a) + R^{u}(a)N(a) \tag{6.20}$$

Combining (6.18)–(6.20) leads to (6.17).

The two terms in (6.17) can be readily explained. The first is the rate at which workers get jobs that pay more than a wage w. This is the rate at which workers move up the job ladder and this effect is why the wage distribution tends to increase with experience. However, some fraction of employees were previously non-employed. These workers have a distribution of wages given by $F(w)$ as they have not had any time to work themselves up the job ladder. They will tend to have a worse distribution of wages than those who have been in employment for some time.

Nothing has been said, so far, about the solution to (6.17). One cannot solve differential equations without an initial condition: in this case we need to specify $G(w|0)$. As new workers will have had no time to work themselves up the job ladder, the most natural assumption is that $G(w|0) = F(w)$ (although one might want to modify this assumption to allow for job search prior to entry into the labor market). The following proposition gives us some useful results.

Proposition 6.2

1. *For a close enough to zero, $E(w|a)$ will be an increasing, concave function of a.*

2. *A sufficient condition for $E(w|a)$ to be increasing in a is that $R^{u}(a')$ is non-increasing in a' for all $a' \leq a$.*

3. *A necessary condition for $E(w|a)$ to be decreasing in a is that $R^{u}(a')$ is increasing in a' for some $a' \leq a$.*

4. *A sufficient condition for $E(w|a)$ to be a concave function of a is that $R^{u}(a')$ is constant for all $a' \neq a$.*

Proof. 1. As the initial condition implies that $G(w|0) = F$, (6.17) then implies that $[\partial G(w|a)/\partial a] < 0$ at $a = 0$. So, for a close enough to zero, $E(w|a)$ must be increasing in a. By further differentiation of (6.17), we have

$$\frac{\partial^2 G(w|a)}{\partial a^2} = -[R^{u}(a) + \lambda[1 - F(w)]]\frac{\partial G(w|a)}{\partial a}$$
$$+[F(w) - G(w|a)]\frac{\partial R^{u}(a)}{\partial a} \tag{6.21}$$

Using the initial condition again we have that $[\partial^2 G(w|a)/\partial a^2] > 0$ for a close enough to zero proving concavity.

2. From (6.17) we know that $[\partial G(w|a)/\partial a] < 0$ if $G(w|a) > G^*(w|a)$ where

$$G^*(w|a) = \frac{R^u(a)F(w)}{R^u(a) + \lambda[1 - F(w)]} \tag{6.22}$$

The only way in which we can have $[\partial G(w|a)/\partial a] > 0$ is if $G(w|a) < G^*(w|a)$ for some (w, a). However, we know that $G(w|0) = F(w) > G^*(w|0)$ so this can only happen if $G(w|a)$ cuts $G^*(w|a)$ from above. As $[\partial G(w|a)/\partial a] = 0$ if $G(w|a) = G^*(w|a)$ this can only happen if $G^*(w|a)$ is increasing in a at some point. But, $G^*(w|a)$ is non-increasing in a if $R^u(a)$ is non-increasing in a. This proves part 2 of the proposition.

3. The argument of the previous section also shows that $R^u(a)$ increasing in a is necessary for expected wages to be declining in a.

4. If $R^u(a)$ is constant then (6.21) says that the earnings function will be concave if it is increasing. Part 2 of the proposition guarantees this.

The Relationship Between Wages, Experience and Tenure in a Pure Search Model

Now let us consider introducing an extra conditioning variable, job tenure, into the analysis. Denote by $G(w|a, t)$ the distribution function of wages conditional on experience and job tenure.

Proposition 6.3. *Expected wages are increasing (decreasing) in job tenure as*

$$R^u(a - t)(R^u(a - t) + \lambda) + \frac{\partial R^u(a - t)}{\partial a} > (<) 0 \tag{6.23}$$

If non-employed workers enter employment at a rate λ_u, and the non-employment rate is $u(a)$, the condition in (6.23) can be written as

$$\delta + (\lambda - \lambda_u)u(a - t) > (<) 0 \tag{6.24}$$

Proof. Denote by $m(w, t|a)$ the proportion of the population of experience a in a job paying wage w and with tenure t. What we know about these people is the following:

- they must have been recruited at experience $(a - t)$;
- they must have been recruited at this time either from non-employment or from jobs paying less than w;
- they have not lost their job or found a better paying one in the time t since then.

The number of workers of experience $(a - t)$ who are recruited into a job paying w is given by $f(w)N(a-t)[\lambda G(w \mid a-t) + R^u(a-t)]$ as a fraction $G(w|a-t)$ of workers in employment will accept a job at wage w and there is also a flow of R^uN non-employed workers, all of whom will accept a job at wage w. The separation rate for these workers is $[\delta + \lambda(1 - F(w))]$ so after time t a fraction $\exp\{-[\delta + \lambda(1-F(w))]t\}$ will remain with the firm. So, $m(w,t|a)$ is given by

$$m(w,t \mid a) = f(w)\exp\left[-[\delta + \lambda(1-F(w))]t\right]$$
$$\times N(a-t)[\lambda G(w \mid a-t) + R^u(a-t)] \quad (6.25)$$

Now, the distribution function of wages conditional on experience and tenure is given by

$$G(w \mid a,t) = \frac{\int_{\underline{w}}^{w} m(v,t \mid a)dv}{\int_{\underline{w}}^{\bar{w}} m(v,t \mid a)dv} \quad (6.26)$$

A method for proving the impact of changes in experience and/or job tenure on expected wages is contained in the following lemma.

Lemma 6.1. *A sufficient condition for the expected wage to be increasing (decreasing) in variable x (experience or tenure) is if $d\ln(m)/dx$ is increasing (decreasing) in the wage.*

Proof. From (6.15) we have that a rise in x will raise the expected wage if we have first-order stochastic dominance (although (6.15) refers to experience, it is obvious that it can also be applied to tenure or any other conditioning variable). To work out the effect on G of x, differentiate (6.26) with respect to x (it is actually more convenient to differentiate $\ln(G)$):

$$\frac{\partial \ln G(w \mid a,t)}{\partial x} = \frac{\int_{\underline{w}}^{w} \frac{\partial m(v,t \mid a)}{\partial x} dv}{\int_{\underline{w}}^{w} m(v,t \mid a)dv} - \frac{\int_{\underline{w}}^{\bar{w}} \frac{\partial m(v,t \mid a)}{\partial x} dv}{\int_{\underline{w}}^{\bar{w}} m(v,t \mid a)dv}$$

$$= \frac{\int_{\underline{w}}^{w} \frac{\partial \ln m(v,t \mid a)}{\partial x} m(v,t \mid a)dv}{\int_{\underline{w}}^{w} m(v,t \mid a)dv} - \frac{\int_{\underline{w}}^{\bar{w}} \frac{\partial \ln m(v,t \mid a)}{\partial x} m(v,t \mid a)dv}{\int_{\underline{w}}^{\bar{w}} m(v,t \mid a)dv}$$

$$= E\left(\frac{\partial \ln m(v,t \mid a)}{\partial x} \mid v \le w\right) - E\left(\frac{\partial \ln m(v,t \mid a)}{\partial x}\right) \quad (6.27)$$

We can sign this unambiguously if the partial derivative of $\ln(m)$ with respect to x is monotonic in w. If it is increasing in w, then $\ln(G)$ is decreasing in x which, from (6.15), means that the expected wage is increasing in x. This proves Lemma 6.1.

Now, returning to the proof of Proposition 6.3 and using Lemma 6.1, we have, by taking the logs of (6.25) and differentiating with respect to t:

$$\frac{\partial \ln m(w,t \mid a)}{\partial t} = -[\delta + \lambda(1 - F(w))] - \frac{1}{N(a-t)}\frac{\partial N(a-t)}{\partial a} - \frac{\lambda \dfrac{\partial G(w \mid a-t)}{\partial a} + \dfrac{\partial R^{u}(a-t)}{\partial a}}{\lambda G(w \mid a-t) + R^{u}(a-t)} \tag{6.28}$$

Now, using (6.17) and (6.20), we can write this as

$$\frac{\partial \ln m(w,t \mid a)}{\partial t} = \frac{-[R^{u} + \lambda(1 - F(w))](\lambda G + R^{u}) - \dfrac{\partial R^{u}}{\partial a} + \lambda^2(1-F)G - \lambda(F-G)R^{u}}{\lambda G(w \mid a-t) + R^{u}(a-t)}$$

$$= \frac{-R^{u}(R^{u} + \lambda) - \dfrac{\partial R^{u}}{\partial a}}{\lambda G(w \mid a-t) + R^{u}(a-t)} \tag{6.29}$$

As G is increasing in w (by definition), the application of the rule in Lemma 6.1 gives (6.23) and (6.3).

To prove the second part of Proposition 6.3, note that, if non-employed workers enter employment at a rate λ_u, then we have

$$R^{u}(a) = \frac{\lambda_u u(a)}{1 - u(a)} \tag{6.30}$$

and

$$\frac{\partial u(a)}{\partial a} = \delta[1 - u(a)] - \lambda_u u(a) \tag{6.31}$$

Differentiating (6.30), and using (6.31) in (6.29) leads to (6.24).

The Distribution of Wages With "True" Returns to Experience and Tenure

In this case, we modify the model of the previous section in the following way. First, we assume that the wage offer distribution changes with experience so that the wage offer distribution at experience a can be written as $F(w - \beta_a(a))$ where $\beta_a(a)$ is a measure of the true returns to experience. We normalize $\beta_a(a)$ so that $\beta_a(0) = 0$. We also assume that wages in a job rise at a rate β_t with tenure. The assumption that the returns to tenure are linear is restrictive but it helps to make the

analysis tractable. Its advantage is that expected wage growth on two jobs is independent of the current wage and tenure so the choice between them can be made solely on the basis of which job offers the highest wage.

As before, we are interested in the cross-section distribution of wages conditional on experience and tenure $g(w|a,t)$: this is derived in the following proposition.

Proposition 6.4

1. *The distribution of wages conditional on experience is given by*

$$[1-u(a)]G(w \mid a) = \exp\left[-\int_0^a [\delta + \lambda(1 - F(w - \beta_a(a) - \beta_t(a-x)))]dx\right]$$
$$\times F(w - \beta_a(a) - \beta_t a)[1-u(0)]$$
$$+\lambda \int_0^a \exp\left[\int_s^a [\delta + \lambda(1 - F(w - \beta_a(a) - \beta_t(a-x)))]dx\right]$$
$$\times F(w - \beta_a(a) - \beta_t(a-s))u(s)ds \qquad (6.32)$$

2. *The distribution of wages conditional on experience and tenure is given by (6.26) but where $m(w,t|a)$ is now given by*

$$m(w,t|a) = \exp\left[-\int_0^t [\delta + \lambda(1 - F(w - \beta_a(a) - \beta_t(t-x)))]dx\right]$$
$$\times f(w - \beta_a(a) - \beta_t t)R \qquad (6.33)$$

where

$$R = [\lambda G(w - \beta_a(a) + \beta_a(a-t) - \beta_t t, a-t)[1 - u(a-t)] + \lambda u(a-t)]$$

3. *The expected level of wages at (a,t) given the returns to experience $\beta_a(a)$ and the return to tenure β_t, $E(w \mid a,t,\beta_a,\beta_t)$ is given by*

$$E(w \mid a,t,\beta_a,\beta_t) = E(w \mid a,t,0,\beta_t) + \beta_a(a) \qquad (6.34)$$

so that one can, without loss of generality, restrict attention to the analysis of earnings functions in which there are no true returns to experience.

Proof

1. As before let us denote by $M(w|a)$ the number of workers of experience a with a wage less than or equal to w. One can think of this group as

being made up of workers who entered employment at different experience levels.

Consider those who were in employment at experience 0. To have wages less than w at experience a, they must have had wages at that time less than $(w - \beta_a(a) - \beta_t a)$ or else the process of wage growth would have made them end up with a wage above w at experience a. There must have been $F(w - \beta_a(a) - \beta_t a)[1 - u(0)]$ of these workers. At date x (where $a \geq x \geq 0$) their wage must have been less than $[w - \beta_a(a) + \beta_a(x) - \beta_t(a - x)]$ so that their instantaneous exit rate from the group will be $[\delta + \lambda(1 - F(w - \beta_a(a) - \beta_t(a - x)))]$. Hence, the exponential part of the first term on the right-hand side of (6.32) gives the fraction of those in employment at experience 0 with wages less than $(w - \beta_a(a) - \beta_t a)$ who are still in employment at a wage less than w at experience a.

Now consider those who last entered employment at experience s. The flow of exits from non-employment at experience s is $\lambda u(s)$. If they have a wage less than w at experience a, they must have had a wage less than $[w - \beta_a(a) + \beta_a(s) - \beta_t(a - s)]$ when they entered employment. This will have been true of a fraction $F(w - \beta_a(a) - \beta_t(a - s))$ of them. They must subsequently not have left employment for non-employment or got a wage offer which means they have a wage higher than w at experience a. At date x (where $a \geq x \geq s$) their wage must have been less than $[w - \beta_a(a) + \beta_a(x) - \beta_t(a - x)]$ so that their instantaneous exit rate from the group will be $[\delta + \lambda(1 - F(w - \beta_a(a) - \beta_t(a - x)))]$. Integrating over all the possible dates of entry to employment gives the second term in (6.32).

2. As we have done before let us denote by $m(w, t|a)$ the density of workers of experience a paid a wage w and with job tenure t. We know that these workers must have been recruited at experience $(a - t)$ at a wage equal to $[w - \beta_a(a) + \beta_a(a - t) - \beta_t t]$. These workers have come from either non-employment or from the mass of workers paid less than this wage at experience $(a - t)$. Hence, the density of recruits is given by

$$f(w - \beta_a(a) - \beta_t t)[\lambda M(w - \beta_a(a) + \beta_a(a - t) - \beta_t t | a - t) + \lambda u(a - t)] \tag{6.35}$$

But, of course, these recruits are only still at this wage at tenure t if they have not left the firm at a tenure $x \in [0, t]$. At tenure x their wage is given by $[w - \beta_a(a) + \beta_a(a - t + x) - \beta_t(t - x)]$. Hence, the proportion of wage offers sufficiently good to make them leave is $[1 - F(w - \beta_a(a) - \beta_t(t - x))]$. Putting this information together, we have

$$m(w, t|a) = \exp\left[- \int_0^t [\delta + \lambda[1 - F(w - \beta_a(a) - \beta_t(t - x))]]dx \right] f(w - \beta_a(a) - \beta_t t)$$
$$\times [\lambda M(w - \beta_a(a) - \beta_t t | a - t) + \lambda u(a - t)] \tag{6.36}$$

Changing the variable of integration in (6.36) from x to $s = (x + a - t)$ leads to (6.34).

3. A simple change of variable from w to $w - \beta_a(a)$ shows that $E(w - \beta_a(a) \mid a, t)$ is independent of $\beta_a(a)$ which gives the result in (6.34).

Let us consider in more detail the likely direction of the bias in the estimated return to tenure. From Proposition 6.3 we know that when $\beta_t = 0$ we only get an unbiased estimate of the return to tenure in the cross-section when the condition in (6.24) is satisfied. The following proposition analyzes one of these cases and shows that the return to tenure will no longer be unbiased and that the sign of the bias depends on the functional form of the wage offer distribution so that it is hard to form any a priori view about it.

Proposition 6.5. *If $u(0) = 0$ and $\delta_u = 0$, the cross-sectional estimate of the earnings function gives a biased estimate of the returns to tenure if $\beta_t > 0$. The bias is positive if*

$$-\lambda \int_0^a [1 - F(w - \beta_t s)]ds + \ln(F(w - \beta_t a)) \qquad (6.37)$$

is a convex function of w for all w and negative if it is a concave function everywhere.

Proof. Let us simplify the algebra by assuming that $\beta_a(a) = 0$. If $u(0) = 0$ and $\delta_u = 0$, no workers are ever in unemployment so that all recruits must come from employment. Using (6.36), we have

$$\begin{aligned} m(w, t \mid a) &= \lambda f(w - \beta_t t) \exp\left[-\lambda \int_0^t [1 - F(w - \beta_t(t - x))]dx\right] \\ &\quad \times G(w - \beta_t t \mid a - t) \\ &= \lambda f(w - \beta_t t) \exp\left[-\lambda \int_0^t [1 - F(w - \beta_t(t - x))]dx\right] \\ &\quad \times \exp\left[-\lambda \int_0^{a-t} [1 - F(w - \beta_t(a - x))]dx\right] F(w - \beta_t a) \quad (6.38) \end{aligned}$$

where the equality follows from the application of (6.32) with the special assumptions made here. Changing the variable of integration to $s = (t - x)$ in the first integral and $s = (a - x)$ in the second, the final line of (6.38) can be written as

$$m(w,t \mid a) = \lambda f(w - \beta_t t) \exp\left[-\lambda \int_0^a [1 - F(w - \beta_t s)] ds\right] F(w - \beta_t a)$$

$$\equiv f(w - \beta_t t)\psi(w, a) \tag{6.39}$$

where $\psi(w, a)$ is the term involving a in (6.39). Application of (6.26) then shows that

$$E(w \mid a, t) = \frac{\int w f(w - \beta_t t)\psi(w, a) dw}{\int f(w - \beta_t t)\psi(w, a) dw} \tag{6.40}$$

Transforming the variable of integration to $v = w - \beta_t t$, we have

$$E(w \mid a, t) = \beta_t t + \frac{\int v f(v)\psi(v + \beta_t t, a) dv}{\int f(v)\psi(v + \beta_t t, a) dv} \tag{6.41}$$

Differentiating this with respect to t, we have

$$\frac{\partial E(w \mid a, t)}{\partial t}$$

$$= \beta_t + \beta_t\left[\frac{\int v f(v)\psi'(v + \beta_t t, a) dv}{\int f(v)\psi(v + \beta_t t, a) dv} - \frac{\int f(v)\psi'(v + \beta_t t, a) dv}{\int f(v)\psi(v + \beta_t t, a) dv}\,\frac{\int v f(v)\psi(v + \beta_t t, a) dv}{\int f(v)\psi(v + \beta_t t, a) dv}\right]$$

$$= \beta_t + \beta_t\left[E\left(v\frac{\psi'}{\psi}\right) - E(v)E\left(\frac{\psi'}{\psi}\right)\right]$$

$$= \beta_t + \beta_t \mathrm{Cov}\left(v, \frac{\psi'}{\psi}\right) \tag{6.42}$$

where the expectation is taken with respect to $(f\psi)$ and ψ' represents the derivative of $\psi(w, a)$ with respect to w. The sign of the bias then depends on the covariance term which depends on whether $\ln(\psi)$ is convex or concave in w. Using (6.39) this can be written as (6.37).

Appendix 6B

Proof of Proposition 6

Although our ultimate interest is simply the unconditional correlation between earnings and experience, we condition on job tenure in what follows because it has previously been shown to be important. One could modify the decomposition to condition on other variables thought relevant.

First, let us introduce the following (lengthy) notation:

- $w(a)$ is the average log wage for workers of experience a;
- $w(a,t)$ is the average log wage for workers of experience a and job tenure t;
- $w^s(a,t)$ is the average log wage for workers of experience a and job tenure t who stay in their jobs over the coming year;
- $w^m(a,t)$ is the average log wage for workers of experience a and job tenure t who move jobs over the coming year;
- $w^l(a,t)$ is the average log wage for workers of experience a and job tenure t who lose their jobs over the coming year;
- $w^e(a)$ is the average log wage for workers of experience a who were not in employment last year;
- $g^s(a,t)$ is the average log wage growth for workers with experience a and job tenure t who stay in their current job;
- $g^m(a,t)$ is the average log wage growth for workers with experience a and job tenure t who move from their current job;
- $n(a)$ is the number of workers of experience a;
- $n(a,t)$ is the number of workers of experience a and job tenure t;
- $n^e(a)$ is the number of workers of experience a who were not in employment last year;
- $\delta(a,t)$ is the fraction of workers of experience a and job tenure t who lose their jobs in the coming year;
- $\lambda(a,t)$ is the fraction of workers of experience a and job tenure t who change their jobs in the coming year.

To keep matters relatively simple, we will restrict attention to a steady state in which all the above are constant over time.

The following relationships must be true by definition.

The current average log wage for those of experience a and job tenure t must be equal to a weighted average of wages for stayers, movers, and job losers:

$$w(a,t) = [1 - \delta(a,t) - \lambda(a,t)]w^s(a,t) + \lambda(a,t)w^m(a,t) + \delta(a,t)w^l(a,t) \tag{6.43}$$

The current average log wage for those of experience a must be equal to a weighted average of the log average wage for those with experience a and different tenure levels:

$$n(a)w(a) = \sum_{t=0}^{a} n(a,t)w(a,t) \tag{6.44}$$

The numbers currently in employment with experience $(a+1)$ and job tenure $(t+1)$ must be equal to those in employment with experience a and job tenure t who have stayed in their jobs:

$$n(a+1,t+1) = [1-\delta(a,t)-\lambda(a,t)]n(a,t) \tag{6.45}$$

The numbers currently in employment with experience $(a + 1)$ must be equal to those in employment with experience a who have not lost their jobs plus entrants to employment at experience $(a + 1)$:

$$n(a+1) = \sum_{t=0}^{a} n(a,t)[1-\delta(a,t)] + n^{e}(a+1) \tag{6.46}$$

Total wages paid out to workers of experience $(a + 1)$ must be equal to a weighted average of the wages for stayers with experience a plus their average growth, the wages for movers with experience a plus their average growth, and the average wage for entrants to employment:

$$n(a+1)w(a+1) = \sum_{t=0}^{a} n(a,t)[[1-\delta(a,t)-\lambda(a,t)][w^{s}(a,t)+g^{s}(a,t)]$$
$$+\lambda(a,t)[w^{m}(a,t)+g^{m}(a,t)]] + n^{e}(a+1)w^{e}(a+1) \tag{6.47}$$

Now let us consider how we can manipulate these identities to end up with (6.13). First use (6.43) to eliminate $w^{s}(a,t)$ from (6.47) which leads to

$$n(a+1)w(a+1)$$
$$= \sum_{t=0}^{a} n(a,t)\Big[w(a,t)+g^{s}(a,t)+\lambda(a,t)\Delta^{m}(a,t)-\delta(a,t)\Big[w^{l}(a,t)+g^{s}(a,t)\Big]\Big]$$
$$+ n^{e}(a+1)w^{e}(a+1) \tag{6.48}$$

where $\Delta^{m}(a,t) \equiv [g^{m}(a,t)-g^{s}(a,t)]$ is an uncorrected measure of the return to job mobility for workers with experience a and job tenure t. As $n(a) = \sum_{t} n(a,t)$, (6.48) can be written as

$$n(a+1)w(a+1) = n(a)\Big[w(a)+\overline{\Delta}^{s}(a)+\overline{\lambda}(a)\overline{\Delta}^{m}(a)\Big] + n^{e}(a+1)w^{e}(a+1)$$
$$-\sum_{t=0}^{a} n(a,t)\delta(a,t)\Big[w^{l}(a,t)+g^{s}(a,t)\Big] \tag{6.49}$$

where $\overline{\Delta}^{s}(a)$ is the average wage growth for stayers, $\overline{\lambda}(a)$ is the average job mobility rate and $\overline{\Delta}^{m}(a)$ is the move-weighted return to job mobility. Using (6.46), we can write (6.49) as

$$n(a+1)w(a+1) = n(a+1)\left[w(a) + \overline{\Delta}^{s}(a)\right] + n(a)\overline{\lambda}(a)\overline{\Delta}^{m}(a)$$
$$+ n^{e}(a+1)\left[w^{e}(a+1) - w(a) - \overline{\Delta}^{s}(a)\right]$$
$$- \sum_{t=0}^{a} n(a,t)\delta(a,t)\left[w^{l}(a,t) + g^{s}(a,t) - w(a) - \overline{\Delta}^{s}(a)\right] \tag{6.50}$$

Now let us define the cost of job loss $\overline{\Delta}^{l}(a+1)$ by

$$\overline{\Delta}^{l}(a+1) = w^{e}(a+1) - \frac{\sum_{t=0}^{a} n(a,t)\delta(a,t)\left[w^{l}(a,t) + g^{s}(a,t)\right]}{\sum_{t=0}^{a} n(a,t)\delta(a,t)} \tag{6.51}$$

This definition compares the earnings of those entering employment from non-employment with an estimate of the earnings those losing their jobs the previous year could have expected if they had stayed in their jobs. If all workers were always in employment, one could compute a more disaggregated measure of the cost of job loss but we typically cannot do that for our data sets. Also, let us define the leaver bias, $b^{l}(a,t)$ as

$$b^{l}(a+1) = w^{l}(a,t) - w(a,t) \tag{6.52}$$

Then, using (6.51) and (6.52), we can write the second and third lines of (6.50) as

$$n^{e}(a+1)\overline{\Delta}^{l}(a+1) + \left[n^{e}(a+1) - \sum_{t=0}^{a} n(a,t)\delta(a,t)\right]$$
$$\times \left[\frac{\sum_{t=0}^{a} n(a,t)\delta(a,t)\left(b^{l}(a,t) + w(a,t) + g^{s}(a,t) - w(a) - \overline{\Delta}^{s}(a)\right)}{\sum_{t=0}^{a} n(a,t)\delta(a,t)}\right] \tag{6.53}$$

The terms in square brackets consist of two sorts of leaver bias. First, the fact that leavers tend to have lower wages than movers and stayers for a given level of experience and job tenure ($b^{l} > 0$). Second, the distribution of tenure among leavers may differ from that in the population as a whole. But the leaver bias only has an effect to the extent that there are changes in employment. Denote the whole leaver bias by $b^{l}(a)$. Putting (6.53) into (6.50) leads to the expression in (6.13).

7

Gender Discrimination in Labor Markets

LABOR market discrimination is usually defined as a situation where workers who are identical in ability have different labor market outcomes. It should not come as a surprise that monopsony or oligopsony has something to say on these issues as, in such a labor market, we know that wages are determined by factors other than productivity. But, this wage dispersion is not quite what is commonly meant by discrimination, a phrase that is generally reserved for a situation where certain groups, for example, women or ethnic minorities systematically receive worse treatment from the labor market.

While the disadvantages suffered by these two groups do have some common features, there are also important differences. For example, the constraints imposed on women by the traditional allocation of domestic responsibilities are not faced by black males, while women do not face the geographical concentration (the extreme form of which is ghettoes, a model of which was presented in section 3.6) that is a feature of the economic predicament of many black workers. This chapter restricts its attention to gender discrimination not because racial discrimination is less important but because monopsony has more to say about gender discrimination.

The outline of the chapter is as follows. The next section documents the most important features of the gender pay gap in the United Kingdom and the United States and shows how these are mirrored in a gender gap in the pattern of labor market transitions, something that is in line with the predictions of a view of the labor market in which employers have some market power. Gender differences in constraints on job search and the reasons for job mobility are then discussed. It is shown that women's job mobility is more constrained by domestic responsibilities than is the case for men and that their job moves are less motivated by money. This shows up in a gender gap in the returns to job mobility. All of this evidence suggests that the wage elasticity in the supply of female labor to a firm should be lower than the male, making the female market less competitive. However, as shown in section 7.6 there is no strong evidence that the wage elasticity of separations is lower for women than for men.

An emphasis on gender differences in labor market attachment is hardly new in discussions of the gender pay gap. The human capital approach also emphasizes these factors but suggests they act to reduce the productivity of women. A direct test of the monopsony and human capital approaches would investigate the role of productivity differences in explaining wage differences: unfortunately direct data on productivity is very rare making it hard to distinguish between the two theories. However, sections 7.8 and 7.9 present two pieces of evidence to suggest that the monopsony approach may be preferable. First, the returns to job tenure appear to be higher for women than men, a finding that is argued to be readily explainable by the monopsony model but more difficult to explain using the human capital approach. Second, an analysis of the impact of the UK Equal Pay Act suggests that the gender pay gap was reduced very substantially without any adverse effect on job opportunities for women.

The chapter concludes with a brief discussion of how the existence of frictions in the labor market is likely to amplify the consequences of any prejudice among employers as proposed by Becker (1971).

7.1 The Gender Pay Gap

Women earn substantially less than men. In the United States, the raw gap in log hourly earnings in 1998–99 was about 19 log points (from the monthly CPS out-going rotation groups): in the United Kingdom, it was about 30 log points (from the LFS). In both countries, the gender pay gap has narrowed in the last 15 years and there was also a rapid narrowing in the United Kingdom from 1970 to 1975 connected with the introduction of the Equal Pay Act. The gender pay gap also has some demographic variation. It is hard to be precise about these variations in the gender pay gap because their existence and size seem to depend on the other variables included in the earnings equation.[1] But, the most consistent variations in the gender pay gap are:

- The return to potential experience (years since left full-time education) is much lower for women than men (although the returns are much closer if actual rather than potential experience is used). There is no earnings gap between men and women in the first 5 years after leaving full-time education in the United Kingdom although a small gap remains in the United States.

[1] Controls for actual experience, job tenure, industry, and occupation seem to be the crucial variables here. For example, Groshen (1991b) finds little gender pay gap exists within occupations within establishments suggesting it is the process of sorting of female workers into jobs that is important in understanding the gender pay gap. However, see Bayard et al. (1999) for rather different results.

- The returns to job tenure are, if anything, somewhat higher for women than men (see, e.g., Becker and Lindsey 1994).
- The pay gap is larger for those who are married and those who have children (the so-called "family penalty" (see Waldfogel 1998a,b)).
- The pay gap is smaller for those from ethnic minorities.

A good explanation of the gender pay gap should be able to explain not just the raw differential but also the variation in the differential. The next section considers how monopsony may explain the gender pay gap.

7.2 Monopsony and the Gender Pay Gap

There is an established "human capital" approach to explaining the gender pay gap. In a perfectly competitive labor market, differences in wages reflect differences in productivity (abstracting from compensating differentials), so the gender pay gap can only be explained by a gender productivity gap. The origin of this gender productivity gap is then identified as being women's weaker attachment to the labor market, mainly the result of the traditional allocation of domestic and child-care labor.

The original discussion of monopsony in Robinson (1933: 302–4) contains an application to the gender pay gap. But, her argument as to why monopsony might be relevant is confined to an example in which men are unionized and women are not, and the slightly enigmatic statement that "a cursory view of existing conditions seems to suggest that [this analysis of the rate of exploitation] may have some bearing upon actual cases" (Robinson 1933: 303). But, monopsony has more to offer on this subject than is apparent from Robinson's analysis. At the end of chapter 2, we emphasized how the monopsonistic approach to labor markets adds other factors to the list of possible determinants of wages, notably labor market transition rates (job offer arrival rates and job destruction rates) and the reservation wage. Even if the productivity of men and women is the same, there will be a gender pay gap if their labor market transition rates differ.

So, a natural starting point for considering how well monopsony can explain the gender pay gap is the discussion at the end of chapter 2. There we proposed a simple statistic, the proportion of recruits from non-employment, as a measure of monopsony power. We have already seen in table 2.2 how a higher fraction of female recruits come from non-employment so that this can explain the overall gender pay gap. But, can it explain the variation in it? Table 7.1 investigates this.

TABLE 7.1
Aspects of the Gender Pay Gap

	US (CPS)		UK (LFS)	
	Log Wage	*Recruited from Employment*	*Log Wage*	*Recruited from Employment*
Single male, children	0.014	−0.027	−0.105	0.024
	(0.004)	(0.008)	(0.008)	(0.009)
Single female, no children	−0.098	0.053	0.029	0.136
	(0.024)	(0.048)	(0.010)	(0.011)
Single female, children	−0.147	−0.031	−0.104	0.017
	(0.024)	(0.048)	(0.009)	(0.011)
Married male, no children	0.114	0.128	0.163	0.176
	(0.004)	(0.009)	(0.005)	(0.009)
Married male, children	0.129	0.111	0.189	0.156
	(0.003)	(0.008)	(0.005)	(0.009)
Married female, no children	−0.071	0.011	0.078	0.183
	(0.024)	(0.048)	(0.010)	(0.012)
Married female, children	−0.093	−0.103	−0.039	−0.015
	(0.024)	(0.046)	(0.010)	(0.013)
R^2	0.38	–	0.4	–
Number of observations	195274	65880	214184	61376

Notes.

1. The reference category is a single male without children.
2. The US regression uses data from January 1998 to June 2000. Other controls include experience (interacted with gender), region, qualifications, month of interview, and black (interacted with gender).
3. The UK regression uses data from December 1992 to November 1999. Married includes those who are living together as a couple. Other controls include experience and tenure (interacted with gender), region, qualifications, month of interview, and black (interacted with gender).
4. In the column headed "recruited from employment" the sample is all those in new jobs and the dependent variable is binary taking the value 1 if the individual was previously employed and 0 otherwise. The reported coefficients are the marginal effects. In the column headed "log wage" the reported coefficients are from an earnings function.

Table 7.1 reports the results from a standard earnings equation and a probit equation where the sample is workers in new jobs and the dependent variable is whether the worker was previously employed (as estimated in table 2.2). Recall that the simple model of chapter 2 predicts less monopsony power in labor markets where a high fraction of recruits were previously employed. The reference category is a single male without children so the estimated coefficient in a particular

row is the difference in the outcome between an individual of that type and the reference category. The most striking differences in earnings in both the United States and the United Kingdom are the premium for being married for men[2] and the penalty for having children for women. For the most part, these differences in wages also show up in the fraction recruited from employment. The evidence in table 7.1 is broadly supportive of a monopsony explanation of the gender pay gap based on differences in labor market dynamics. Bowlus (1997) and Barth and Dale-Olsen (1999) take a different approach to the same idea, showing how, in the Burdett and Mortensen (1998) model, the degree of monopsony power is a function of the labor market transition rates and showing how these are systematically different for men and women.

The gender difference in labor market transition rates provides an incentive for employers to pay otherwise identical men and women different wages (for further explanation of this, see Bowlus 1997; Barth and Dale-Olsen 1999). But, it is important to realize that the differences in labor market transition rates will result in a gender pay gap even if individual employers do not discriminate against women. Women will simply find it harder to work their way up the job ladder[3] and a gender pay gap will remain even if there is fully effective equal pay legislation. Groshen (1991b) concluded that the gender pay gap disappears once one includes detailed controls for occupation and employer suggesting that this effect may be more important in current labor markets. However, Bayard et al. (1999) cast some doubt on Groshen's conclusion, finding that there are still substantial gender pay differences within occupations within particular employers. This is an important issue where further research would be welcome.

However, the analysis so far is all based on the very simple version of the Burdett–Mortensen model of chapter 2.4 and there are good reasons for wanting to delve a little deeper as that model is overly simplistic in at least two areas which might be thought to be very relevant for explaining the gender pay gap.

First, the simple Burdett–Mortensen model assumes that labor is supplied inelastically to the market as a whole. Conventional wisdom says that individual labor supply is much more elastic for women than

[2] How these family effects should be interpreted is a matter of some debate. One interpretation would be a causal one from marriage and/or children to earnings. Others would emphasize how these household characteristics are correlated with unobserved factors (perhaps the old standby "ability") that also correlate with earnings.

[3] Proposition 6.1 is relevant here as it shows that an increase in the fraction of recruits from non-employment results in a lower position of workers on the job ladder.

men.[4] Secondly, the simple model assumes that all job moves are motivated by money. There are good reasons for thinking that there may be important gender differences in the reasons for job mobility that may affect the average level of earnings. The next few sections pursue these issues.

7.3 The Elasticity in Labor Supply to the Firm and the Market

There are a number of ways of introducing some elasticity into the supply of labor to the market as a whole into the basic model of an oligopsonistic labor market. Two ways of doing so were introduced in chapter 3: a fixed cost of entry (the model of section 3.3) and heterogeneity in reservation wages (the model of section 3.4). There are others: for example, one could make search intensity endogenous in which case the arrival rate of job offers will depend on the expected return to search. This line is discussed in chapter 9 but it is not the main reason why labor economists think that female labor supply is more elastic. More in line with traditional thinking is to introduce some heterogeneity in reservation wages.

A model of this type has already been introduced in section 3.4 where it was assumed that there is a distribution of the value attached to leisure, b, in the population. Denote by $H(b)$ the fraction of the population with value of leisure b or less. $H(b)$ tells us about the elasticity of the labor supply curve to the market as a whole. Perhaps the easiest way to see this is to consider a simple monopsonist who would choose the wage to maximize $(p - w)H(w)$. For a single monopsonist, $H(w)$ is both the supply of labor to the individual firm and to the market as a whole.

Why should $H(w)$ be more elastic for women than men? One explanation is that there is more heterogeneity in the value attached to home time among women than men because of greater heterogeneity in domestic responsibilities. Another explanation is based on the greater prevalence of part-time work among women combined with fixed costs of going to work. Suppose there is a simple monopsonist who faces an upward-sloping supply curve of labor, $H(U)$, where U is the utility that workers obtain from employment. Assume, for simplicity, that the elasticity of this supply curve is a constant and the same for men and women: denote it by ε_{HU}. Suppose that each hour of work pays w but has disutility b and

[4] Indeed, some economists have argued against Joan Robinson's views on the grounds that female labor is supplied more elastically than male labor. There is something a little odd about this argument as its proponents generally believe in competitive labor markets in which there is an enormous assumed gap between the elasticity in the supply of labor to the market as a whole and to an individual firm (as that is assumed infinite). But, there is clearly an issue here that needs to be sorted out.

that there is a fixed cost C of going to work. Utility is given by $U = (w - b)e - C$ where e is hours of work. The employer will be concerned about the wage elasticity of the labor supply curve which is given by

$$\varepsilon_{Hw} = \frac{w}{H}\frac{\partial H}{\partial w} = \frac{w}{U}\frac{U}{H}\frac{\partial H}{\partial U}\frac{\partial U}{\partial w} = \frac{we}{(w-b)e - C}\varepsilon_{HU} \qquad (7.1)$$

Suppose some individuals ("women") work part-time, that is, e is low. Then, (7.1) says that the wage elasticity of the labor supply curve facing the employer will be higher because a given percentage increase in the wage leads to a higher percentage increase in the utility.

Now, consider what happens to the labor supply curve facing an employer when we introduce other firms into the market. The following proposition provides the answer using the notation introduced in chapters 2 and 3.

Proposition 7.1. *The supply of labor to an individual firm is given by*

$$N(w) = \frac{\delta\lambda H(w)}{M[\delta + \lambda(1 - F(w))]^2} \qquad (7.2)$$

Proof. See proof of Lemma 3.1.

(7.2) shows that the elasticity of the labor supply to an individual firm combines, in a neat way, both the elasticity of the labor supply to the market as a whole (measured by $H(w)$) and the effect of competition from other firms (measured by $F(w)$). Note that as the labor market becomes more competitive (λ increases), the contribution of the supply of labor to the market as a whole becomes relatively less important. In the limiting case of perfect competition, $(\lambda/\delta) \to \infty$, (7.2) says that the supply of labor to the individual employer will be infinitely elastic at the maximum wage offered in the labor market. Using (7.2) to think about reasons for a gender gap in the elasticity of labor supply to the firm, competition from other firms may be less severe in the case of women but the elasticity of supply to the market as a whole is likely to be higher. So, it is unclear, a priori, whether the elasticity in the supply of female labor to a firm is greater or less than the male elasticity.

7.4 Money and Motivation

Another way in which the previous discussion may be too simplistic is that it assumes that all job moves are motivated by money and that a wage gain of a cent is enough to guarantee a move. In reality, other

factors are important, for example, the location of the job, the hours offered, etc. Altonji and Paxson (1988, 1992) use the PSID to show that workers who obtain more favorable hours when they change jobs have lower wage growth suggesting that workers are trading off pecuniary and non-pecuniary aspects of jobs. It is quite likely that these other factors loom larger in the minds of women who often have the twin burdens of domestic responsibility for children and work for money.

The UK BHPS asks some questions on whether domestic responsibilities act as a constraint on job search and mobility. The answers to these questions are tabulated in table 7.2 and show that the extent of constraints on job search and mobility is much greater for married women with children than for either men or single women. For example, only 1.4% of single men report that domestic constraints have prevented job change compared to 13.2% of married women with children.

The UK BHPS also asks individuals about their reasons for changing jobs. The answers are tabulated in table 7.3. More men than women (26.7% against 24.3%) and more single women than married women with kids (22.9% against 20.4%) report that their job move was the result of a promotion, a pecuniary factor. Similar proportions of men and women report moving for a better job, but, as table 7.4 makes clear, many more men (47.1%) than women (33.8%) report pecuniary factors as the reason the job is better. Although there remains a gap between men and women in the motivation for leaving jobs, this gap is much narrower now than it used to be. Part of the BHPS collects retrospective data on life-time employment history including a question on the reason for leaving jobs. Figure 7.1a presents the proportion of job changes that were motivated by "career" concerns by year of job change:[5] men are more likely to be driven by these concerns but the gap is much smaller than it used to be. Figure 7.1b presents a similar picture but for the fraction of job changes motivated by "domestic" concerns: again, we see the gender gap declining.

All this adds up to a picture in which women and men differ in the attributes of jobs that are most valued with men putting relatively more emphasis on money. Similar results are reported for the United States by Keith and McWilliams (1999) for the NLSY and by Sicherman (1996) who uses a firm-level data set to show that, while overall levels of job turnover are similar for men and women, the reasons for turnover are very different with women much more likely than men to leave their jobs for non-market reasons.

[5] There is a problem with identifying life-cycle and cohort effects here but a regression in which there is a control for age comes up with the same conclusions: there is a convergence in motivation between men and women.

TABLE 7.2
Domestic Constraints on Job Search: UK BHPS

	Prevented Job Search	*Prevented Taking Job*	*Prevented Job Change*	*Required Job Change*	*Required Leaving Job*	*Led to Less Work Hours*	*Number of Observations (Approx.)*
All	3.6	2.9	4.4	1.3	0.6	2.9	31704
Men	1.5	0.8	2.3	0.7	0.2	1.2	15761
Women	5.7	5.0	6.4	1.9	1.0	4.7	15943
Married men	1.6	0.8	2.7	0.8	0.2	1.4	11280
Single men	1.3	0.6	1.4	0.2	0.1	0.6	4480
Married women	6.1	7.6	6.9	2.1	1.1	7.3	11509
Single women	4.8	3.6	5.0	1.4	0.8	2.9	4434
Married men, children	2.2	1.2	4.0	1.2	0.3	2.2	5583
Married men, no children	1.0	0.4	1.4	0.5	0.2	0.5	5697
Married women, children	12.3	11.2	13.2	3.3	1.9	10.5	5084
Married women, no children	1.2	1.1	2.0	1.1	0.5	1.2	6425

Notes.
1. The question asked is whether family commitments prevented job search, taking job, etc.
2. The sample is all those in employment. The number of observations differs slightly from column to column. The source is the BHPS 1991–98.

TABLE 7.3
Reasons for Jobs Ending: UK BHPS

	Promoted	*Better Job*	*Lost Job*	*Retirement/ Health*	*Domestic Responsibilities*	*Other*	*Number of Observations*
All	27.6	37.7	16.6	1.8	0.7	19.7	9870
Men	26.7	37.5	17.5	1.6	0.1	18.5	5101
Women	24.3	37.6	17.7	1.9	1.3	20.9	4769
Married men	28.3	33.2	16.3	1.7	0.1	19.9	3427
Single men	22.5	40.3	19.8	1.6	0.1	17.8	1673
Married women	24.8	37.2	14.6	2.1	1.4	21.9	3187
Single women	22.9	38.8	18.9	1.5	1.1	19	1581
Married men, children	29.8	32.8	14.7	1.2	0.1	21.4	1764
Married men, no children	27.8	33.6	18.1	2.2	0.0	18.3	1662
Married women, children	20.4	36.3	17.7	1.9	2.8	23.5	1400
Married women, no children	28.3	34.3	14.1	2.3	0.3	20.7	1787

Notes.
1. The sample is all those who left a job for a new job with another employer. The source is the BHPS 1991–98.

TABLE 7.4
Reasons Given for Why New Job Is Better: UK BHPS

	Pecuniary Factors	*Non-pecuniary Factors*	*Number of Observations*
All	40.6	59.4	2717
Men	47.1	52.9	1389
Women	33.8	66.2	1328
Married men	47.8	54.1	892
Single men	49.4	50.6	496
Married women	30.7	69.3	892
Single women	40.0	60.0	435
Married men, children	47.2	54.8	449
Married men, no children	46.5	53.5	443
Married women, children	29.3	70.7	396
Married women, no children	31.8	68.1	496

Notes.
1. The sample is all those who left a job for a new job with another employer and reported that the reason they left the previous job was because they had obtained a better job. The source is the BHPS 1991–98.

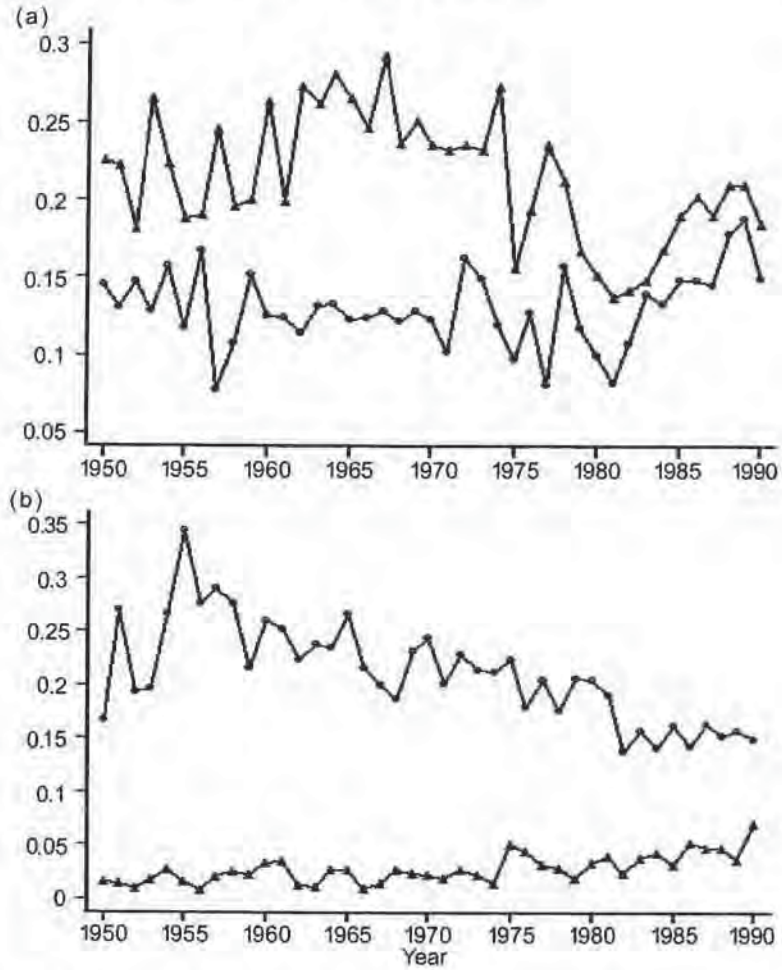

Figure 7.1 (a) The fraction of jobs ending with move to a better job. ○, men; △, women. (b) The fraction of jobs ending because of domestic responsibilities. *Notes.* Both (a) and (b) are from the lifetime history record collected in the BHPS in 1994. (a) is the fraction of jobs ending where the reason given was "better job" while (b) presents the fraction giving "to have a baby" or "domestic responsibilities" as the reason. The year is the year in which the job ended.

Another aspect of the constraints on women is in terms of how far they can travel to work. If women are constrained by domestic responsibilities to take part-time work, then the fixed time and money costs of getting to work become relatively more important so that the worker is likely to be less prepared to travel long distances to work thus restricting the range of

TABLE 7.5
Travel-to-Work Times in Britain

	(1) All	(2) Men	(3) Women	(4) Men	(5) Women	(6) Men	(7) Women
Constant	27.11 (0.06)	27.21 (0.16)	23.20 (0.11)	27.00 (0.16)	24.33 (0.11)	–	–
Female	−6.12 (0.09)						
Married		2.62 (0.18)	−3.20 (0.13)	2.16 (0.19)	−2.87 (0.13)	1.35 (0.22)	−0.72 (0.13)
Number of children				0.66 (0.07)	−1.74 (0.05)	0.26 (0.07)	−2.37 (0.05)
Other controls	No	No	No	No	No	Yes	Yes
Number of observations	260504	123313	114915	123059	114797	121939	113955
R^2	0.018	0.002	0.006	0.002	0.015	0.561	0.619

Notes.
1. The dependent variable is the usual one-way home to work travel time in minutes. Sample is from the autumn UK LFS 1995–99.
2. Standard errors in parentheses.
3. Married is a 0–1 dummy variable taking the value 1 if the individual is married or living as one of a couple. Number of children is the number of dependent children under the age of 19 in the household.
4. The other controls are educational qualifications, race, county of residence, the year and month, and a quadratic in experience.

possible jobs. Table 7.5 presents some evidence on travel-to-work times from the UK LFS. On average, men take 27 minutes to get to work while women take 6 minutes less (column (1)). Married men travel more than single men (column (2)) while married women travel less (column (3)). Men with children travel more than childless men while the reverse is true for women (columns (4) and (5)). These effects survive the introduction of controls (columns (6) and (7)). This suggests that women, particularly women with domestic responsibilities, are restricted in the distance they can travel to work, one of the non-pecuniary factors that is likely to lower their earnings and make their labor market less competitive.

The bottom line from this discussion is that there are reasons to believe that female job-to-job changes are less sensitive to the wage than male and that the wage gains from job mobility are lower for women. The next two sections investigate these predictions.

7.5 Gender Differences in the Returns to Job Mobility

This section explores whether there are any significant gender differences in the returns to job mobility. The qualitative evidence presented in the previous section suggested that we would expect to see significantly lower returns from job mobility for women than for men.

Table 7.6 presents estimates of the returns to job mobility for men and women using our three panel data sets, the US PSID and NLSY, and the UK BHPS. The dependent variable is the change in log wages for workers in continuous employment. We include as controls a dummy variable for gender and a gender dummy interacted with whether the individual has changed jobs.

For each data set we present two specifications, one in which there are no controls apart from the dummy variables for gender and having moved jobs plus the interaction between them, and one in which other controls are included. For the PSID, the gender gap in the return to job mobility (the coefficient on the female mover variable) is significantly negative so that the returns to job mobility are lower for women than for men, whether controls are included or not. For the BHPS, we obtain the same result, although the gap is only significant once other controls are included in the equation. The

TABLE 7.6
Gender Differences in the Returns to Job Mobility

	PSID		*NLSY*		*BHPS*	
Mover	0.0586	0.0812	0.0120	0.0246	0.0539	0.1031
	(0.0058)	(0.0087)	(0.0080)	(0.0119)	(0.0097)	(0.0133)
Female mover	−0.0158	−0.0181	0.0125	0.0091	−0.0214	−0.0310
	(0.0083)	(0.0082)	(0.0125)	(0.0125)	(0.0134)	(0.0134)
Female	0.0051	0.0027	−0.0044	−0.0065	−0.0014	−0.0019
	(0.0023)	(0.0023)	(0.0065)	(0.0065)	(0.0041)	(0.0042)
Other controls	No	Yes	No	Yes	No	Yes
Number of observations	53053	53052	16030	15971	17800	17699
R^2	0.0030	0.0116	0.0005	0.0053	0.0025	0.0201

Notes.
1. The dependent variable is the change in log hourly wages. The sample is those workers in continuous employment from one year to the next.
2. Where included, the other controls are mover status interacted with lagged experience and job tenure, a quadratic in lagged experience, lagged job tenure, education, ethnicity, region, and year dummies.

NLSY has rather different results with no evidence of lower returns to job mobility for women.

However, there are a number of other studies which have examined gender differences in the returns to job mobility using the NLSY. Loprest (1992) reports significantly lower returns to job mobility for women at the very start of their labor market careers. Keith and McWilliams (1997) extended this analysis by disaggregating the causes of separation. They concluded that the returns to different types of job mobility were similar for men and women but that there was a gender gap in the incidence of the different types of job mobility. Women were more likely to have a family-related quit (which has a wage penalty) and less likely to have voluntary job-to-job separations. Keith and McWilliams (1999) extended this analysis to investigate the returns to job-related search finding, again, the returns to this activity were similar for men and women but women do less of it

7.6 Gender Differences in the Wage Elasticity of Separations

The previous sections have highlighted gender differences in the labor market transitions of men and women. A consistent picture in which women's job opportunities are more constrained and job decisions less motivated by money has emerged. It is well known that the separation rate is higher for women than men (although the gap is much less than it used to be) but these factors suggest that we might expect to find a gender gap in the wage elasticity of separations. This section examines this prediction.

Table 7.7 presents estimates of separation elasticities for the PSID and NLSY for the United States, and for the BHPS and LFS for the United Kingdom. These estimates are obtained using the approach of section 4.7. The reported coefficients measure the elasticity of the separation rate with respect to the wage. We show the estimated elasticities for two specifications: both without and with the controls listed at the bottom of the table.

First, consider the wage elasticity for all separations. Without controls, the male elasticity is higher than the female for three of the four data sets although the differences are small and not significantly different from zero. However, this result does not stand up to the introduction of controls: now the female elasticity is larger than the male in three of the data sets and the gap is actually more significant. We might expect female separations to non-employment to be more sensitive to the wage than male separations. Without controls, there is no evidence for this although this pattern does emerge in three of the four data sets once

TABLE 7.7
The Wage Elasticity of Separations with Respect to the Wage

Sample	*Gender*	*No controls*			*Controls*		
		All Separations	*Separations to Other Jobs*	*Separations to Non-employment*	*All Separations*	*Separations to Other Jobs*	*Separations to Non-employment*
PSID	Male	−1.005	−0.927	−1.046	−0.880	−0.889	−0.868
		(0.055)	(0.054)	(0.088)	(0.058)	(0.038)	(0.085)
PSID	Female	−0.971	−0.744	−1.059	−1.055	−0.936	−1.101
		(0.034)	(0.042)	(0.039)	(0.045)	(0.055)	(0.048)
NLSY	Male	−0.580	−0.500	−0.676	−0.554	−0.544	−0.507
		(0.042)	(0.056)	(0.049)	(0.046)	(0.062)	(0.057)
NLSY	Female	−0.548	−0.415	−0.651	−0.629	−0.575	−0.678
		(0.041)	(0.066)	(0.043)	(0.048)	(0.079)	(0.049)
BHPS	Male	−0.968	−0.914	−1.042	−0.742	−0.753	−0.735
		(0.058)	(0.079)	(0.081)	(0.078)	(0.107)	(0.110)
BHPS	Female	−0.901	−0.886	−0.917	−0.566	−0.471	−0.677
		(0.091)	(0.131)	(0.118)	(0.120)	(0.172)	(0.162)
LFS	Male	−0.642	−0.591	−0.726	−0.452	−0.481	−0.414
		(0.029)	(0.039)	(0.042)	(0.042)	(0.053)	(0.068)
LFS	Female	−0.652	−0.565	−0.744	−0.540	−0.438	−0.659
		(0.036)	(0.047)	(0.044)	(0.039)	(0.057)	(0.055)

Notes.
1. The samples are the same as those used in table 4.7.
2. All equations contain the following additional controls: education, race, marital status, children, region, a quartic in experience, and year dummies. Tenure is excluded: its inclusion lowers the estimated wage elasticities while preserving the qualitative results. See section 4.7 for a discussion of whether tenure should be included or excluded from these equations.

controls are included. Conversely, one might expect that the male elasticity is greater than the female elasticity for separations to other jobs. There is weak evidence for this although, again, the gender gap is not very large or robust.

The results presented here are consistent with other estimates of these elasticities. Viscusi (1980), using data from the PSID in 1975–76 found a wage elasticity close to −1.0 for both men and women. Other estimates can be found in Royalty (1998) where, although the specification estimated does not make it very easy to work out the implied elasticities, there does not seem to be a very striking gender gap.

One possible explanation for the finding that there is no systematic gender gap in the wage elasticity of separations is that the differences in the labor turnover behavior of men and women are now quite small. The differences in separation rates were almost certainly once higher, for example, Viscusi (1980) reports that female quit rates in US manufacturing were 80% above the male rate in 1958, but only 16% higher in 1968. However, the finding of Viscusi (1980) that there was no gender gap in wage elasticities in the mid-1970s casts some doubt on the hypothesis that female wage elasticities used to be below those of men.

Consequently, the gender differences that we have identified in previous sections do not show up strongly in these estimated elasticities. Whether this is because this approach to estimating elasticities is not very informative or because the total effect of the gender differences in constraints and motivation is small, is an issue that deserves further consideration.

7.7 Human Capital Explanations of the Gender Pay Gap

The discussion so far has paid little or no attention to well-established explanations of the gender pay gap based on human capital theory. These theories also emphasize the constraints imposed on women by the traditional division of labor within the household but model the impact on wages as being through an impact on productivity. Perhaps the simplest way to see the human capital approach at work is in the papers that make the distinction between actual and potential experience and investigate whether the gender differences in the returns to potential experience that we have already noted, can be explained in terms of a common return to actual experience, and gender differences in the level of experience (see, e.g., Light and Ureta 1995).[6] Direct data

[6] On the whole this literature claims some success although the search approach can also explain such findings and a substantial gender pay gap remains. And, it is typically found that any interruption in paid employment, however short, reduces earnings.

on productivity are rare so it is difficult to provide a direct test of the human capital and monopsony approaches. The study of Hellerstein et al. (1999) finds evidence that the gender pay gap across firms is much larger than the gender productivity gap suggesting that the human capital approach is not the whole story. But, in the absence of direct data on productivity, there are some differences in implications of the two theories that one might hope to exploit.

First, consider the returns to tenure. In the human capital approach, the returns to tenure are thought to represent a share of the returns to investments in firm-specific human capital. If women are more likely to leave a firm than men, we would then expect the incentives for investments in firm-specific human capital to be reduced and the observed returns to tenure to be lower for women as a given level of job tenure is likely to be associated with a lower level of firm-specific human capital. In contrast, the search model predicts a stronger correlation between tenure and wages for those groups with weak labor market attachment as tenure becomes a better measure of the time elapsed since the individual was last non-employed (see table 6.2 and the discussion surrounding it).[7] Table 7.8 reports estimates of the cross-sectional returns to tenure for women and men from the US CPS and the UK LFS. The reported returns are the estimated gap in earnings between someone with the reported years of tenure and someone just starting a job. For the United States, the female return to job tenure is noticeably *higher* than the male at all experience and tenure levels. In the United Kingdom, the gender gap in the returns to tenure is much smaller but there is not much evidence that the male return to job tenure is higher. This evidence is consistent with the monopsony approach. Such an empirical finding is not new: Becker and Lindsay (1994) report estimates from the US PSID consistent with this. They chose to explain the empirical findings using a variant of the Hashimoto (1981) model in which, although incentives to invest in firm-specific human capital are lower for women, they are predicted to receive a higher share of those returns. However, they obtain this result by assuming that women quitters are harder to identify, an arbitrary assumption that is at variance with the findings of Light and Ureta (1992).

Another approach to distinguishing between monopsony and competitive approaches to the gender pay gap is to look at predicted differences in response to policy changes. This approach is taken in the next section where we investigate the impact of UK equal pay legislation.

[7] Also, see Manning (1998), who compares actual returns to tenure with those predicted from a simple search model and observed labor market transition rates.

TABLE 7.8
Gender Differences in the Returns to Job Tenure

	5 years of job tenure		*10 years of job tenure*		*20 years of job tenure*	
	US CPS	*UK LFS*	*US CPS*	*UK LFS*	*US CPS*	*UK LFS*
10 years of potential experience						
Men	0.129	0.156				
	(0.016)	(0.005)				
Women	0.216	0.187				
	(0.016)	(0.005)				
20 years of potential experience						
Men	0.124	0.168	0.217	0.244		
	(0.018)	(0.005)	(0.020)	(0.006)		
Women	0.200	0.184	0.326	0.294		
	(0.018)	(0.005)	(0.019)	(0.006)		
30 years of potential experience						
Men	0.133	0.146	0.228	0.238	0.350	0.328
	(0.024)	(0.006)	(0.023)	(0.007)	(0.022)	(0.006)
Women	0.161	0.111	0.281	0.200	0.444	0.326
	(0.022)	(0.006)	(0.022)	(0.006)	(0.022)	(0.007)

Notes.

1. The US data come from the 1996 and 1998 Job Tenure Supplements to the CPS. The estimated returns to experience and job tenure come from a quartic in both variables with all interactions between them. Other controls included are: education, black, a year dummy, state dummies, dummies for being married, and the presence of children (and the interaction between them). Total number of observations is 19,281.
2. The UK data come from the LFS from 1992 to 1999. The estimated returns to experience and job tenure come from a quartic in both variables with all interactions between them. Other controls included are: education, black, Asian, month dummies, region dummies, dummies for being married, and the presence of children (and the interaction between them). Total number of observations is 212,478.
3. Standard errors in parentheses.

7.8 The Effect of UK Equal Pay Legislation

In the early 1970s two pieces of legislation designed to attack labor market discrimination against women were passed in the United Kingdom. The Equal Pay Act which essentially required equal pay for men and women doing similar work was passed in 1970 but did not have the force of law until the end of 1975. The 1975 Sex Discrimination Act, which came into force at the same time, made it unlawful to discriminate against women in matters relating to access to jobs. There is little doubt that the

two pieces of legislation had a substantial effect on the relative earnings of women as, having been approximately constant for the period prior to the early 1970s, they then rose sharply and were roughly constant thereafter until they began to rise again in the mid-1980s.[8] This rise in women's relative wages over a short period of time is almost certainly the sharpest change in relative wages in the post-war period and it confronted employers with a change in relative wages that was largely exogenous to them. If labor markets were initially competitive with all workers, male and female, being paid their marginal products one would expect the rise in female relative earnings to be mirrored by a fall in relative employment. This prediction would be shared by other models (perhaps Becker's (1971) model of discrimination) in which the introduction of equal pay legislation would be predicted to lead to a situation in which there is an excess supply of female labor. On the other hand, if labor markets were monopsonistic the rise in women's wages might actually lead to an increase in employment. So, one can think of the experience of the Equal Pay Act as providing helpful insights into the workings of the labor market.

Unfortunately, the period of the introduction of the Equal Pay Act was largely before the availability of micro data. The best data available for investigation of its impact is industry-level data for manufacturing industries: in what follows we use data on something like 100 three-digit industries (for more detailed description of the data, see Manning 1996). From the early 1960s to the late 1960s the wages of women in manufacturing actually fell by 1.3% relative to men, but from the early 1970s to the late 1970s they rose by 21%. In terms of relative employment, female relative employment fell by an average of 7.7% in the 1960s and 2.1% in the 1970s (although the growth of other sectors of the economy meant that relative employment of women rose in the economy as a whole). That the relative employment of women fell more slowly at a time when their relative wages were rising faster is interesting but not persuasive evidence in favor of monopsony.

But, we can exploit the fact that the Equal Pay Act had a different impact on different industries. One of the reasons that the Equal Pay Act had such a quick, large effect on the relative wages of women was the prevalence of collective bargaining (overall union coverage was approximately 75%). Prior to the Equal Pay Act it was commonplace for union-negotiated contracts to have lower wages for women on the same jobs written into them (typically at the "biblical" fraction of two-thirds). Once the Equal

[8] The large impact of equal pay legislation in the United Kingdom is not mirrored in all other countries. In particular, the Equal Pay legislation of 1963 in the United States did not seem to have any very noticeable impact on women's relative pay. We will suggest explanations for this below.

Pay Act was in force, this was clearly illegal and the contractual pay rates of women and men on the same jobs had to be equal. So, we might expect the Equal Pay Act to have raised the relative wages of women more in industries where collective bargaining was more prevalent. Secondly, the Equal Pay Act contained a clause stating that the lowest pay rate for a woman on any job could not be below the lowest pay rate for a man on any job (a crude attempt to deal with the occupational segregation of men and women). So, we might expect to see a larger rise in the relative wages of women where the level of relative wages was initially low.

Both of these hypotheses seem to be confirmed by the data. The first column of table 7.9 looks at the impact on relative wage changes in the 1970s of the initial level of relative wages and the levels of female and male union coverage. High levels of female union coverage are associated with higher increases in relative wages as are lower initial levels of the relative wage. That it is plausible to ascribe these effects to the Equal Pay Act is given additional support by estimating a similar equation for relative wage changes in the 1960s reported in the second column. None of the regressors are significant and relative wages seem to have followed a random walk in this period.

However, we are more interested in the impact of the Equal Pay Act on the relative employment of women if we want to have some insight into the workings of labor markets. To keep things simple, assume that the revenue function can be written as $Y(N, X)$ where N is a composite measure of labor inputs and X is all other inputs. N is assumed to be given by

$$N = \left[AN_f^{\chi} + (1 - A)N_m^{\chi}\right]^{1/\chi} \qquad (7.3)$$

where N_f is female employment and N_m is male employment.

Assume, in the interests of generality that the labor markets for both men and women might be monopsonistic and that the labor supply curve to the firm is given by

$$W_i = B_i N_i^{\varepsilon_i}, \qquad i = f, m \qquad (7.4)$$

In the absence of equal pay legislation we assume that the firm is free to choose W_f independent of W_m. A necessary condition for profit maximization is that (N_m, N_f) (and the associated wages) are chosen to minimize labor costs subject to the constraint that the labor index is at a particular level, N, that is,

$$\min_{(W_f, W_m, N_f, N_m)} W_f N_f + W_m N_m$$

$$\text{s.t.} \quad AN_f^{\chi} + (1 - A)N_m^{\chi} = N^{\chi}, \qquad W_i = B_i N_i^{\varepsilon_i}, \qquad i = f, m \quad (7.5)$$

TABLE 7.9
The Impact of the UK Equal Pay Act

Dependent Variable	*Change in Log Wage of Women Relative to Men*		*Change in Log of Female Employment Divided by Male Employment*			
	1970s	*1960s*	*1960s*	*1970s*	*1970s*	*1960s*
Estimation method	OLS	OLS	OLS	OLS	IV	IV
Initial relative wage	−0.234	−0.038				
	(0.051)	(0.044)				
Female union coverage	0.134	0.001				
	(0.047)	(0.003)				
Male union coverage	−0.016	−0.038				
	(0.039)	(0.029)				
Change in relative wage of women			0.315	−0.099	0.234	0.394
			(0.344)	(0.173)	(0.382)	(2.045)
Constant	−0.002	−0.004	−0.067	−0.0004	−0.069	−0.069
	(0.036)	(0.030)	(0.012)	(0.035)	(0.076)	(0.0.29)
Number of observations	108	96	103	113	107	96
R^2	0.24	0.04	0.01	0		

Notes.

1. The data in this regression are from three-digit manufacturing industries. Data on employment come from the Census of Employment, data on hourly earnings from the October Earnings Enquiry and data on union coverage from the 1973 and 1978 New Earnings Survey. Relative earnings (employment) growth for the 1960s (1970s) refers to growth between the average level of earnings (employment) in 1965–68 (1975–78) compared to the average in the period 1960–62 (1970–72).

Simple manipulation of the first-order conditions of the solution to (7.5) allows us to derive

$$\log\left(\frac{N_f}{N_m}\right) = \frac{1}{1-\chi}\log\left(\frac{A}{1-A}\right) - \frac{1}{1-\chi}\left[\log\left(\frac{W_f}{W_m}\right) + \log\left(\frac{1+\varepsilon_f}{1+\varepsilon_m}\right)\right] \tag{7.6}$$

This simply says that the ratio of the marginal products must be equal to the ratio of the marginal costs of labor. It is more convenient to write (7.6) in difference terms so that we have

$$\Delta\log\left(\frac{N_f}{N_m}\right) = \frac{1}{1-\chi}\Delta\log\left(\frac{A}{1-A}\right) - \frac{1}{1-\chi}\left[\Delta\log\left(\frac{W_f}{W_m}\right) + \Delta\log\left(\frac{1+\varepsilon_f}{1+\varepsilon_m}\right)\right] \tag{7.7}$$

Suppose that the labor market was competitive (set $\varepsilon_f = \varepsilon_m = 0$ in (7.7)). Then, controlling for relative demand shifts, we should expect to see a negative correlation between relative employment and relative wage changes. The third column of table 7.9 looks at this correlation in the 1960s. The correlation between relative employment and relative wage changes is positive although insignificantly different from zero. One plausible explanation for this is that the estimated coefficient is biased because, if the labor supply to an industry is not perfectly elastic (which is still consistent with the labor supply curve to an individual firm being perfectly elastic), there will be a positive correlation between the relative demand shocks $A/(1-A)$ and the relative wage in (7.7).

But, the earlier analysis suggested that relative wage changes in the 1970s were driven in large part by factors other than relative demand shifts. So we might expect to see a more negative correlation between relative employment and relative wage changes in that period. The fourth column of table 7.9 provides some support for this view although the negative correlation between relative wage and relative employment changes is not significantly different from zero. However, we can do better than this. If we want to get a consistent estimate of the wage elasticity of relative demand, then we want to instrument relative wages using variables that are not correlated with the relative demand shocks. Our earlier regressions explaining relative wage changes suggested good instruments (the initial relative wage and the unionization rate) that were correlated with relative wage changes because of the Equal Pay Act. But, the IV results, reported in the fifth column, actually make the estimated elasticity positive, although not significantly different from zero (one can accept the over-identifying restrictions implied by this specification). For completeness the sixth column presents the equivalent

results for the 1960s but the weakness of the instruments for that period shows up in an enormous standard error.

These results are hard to explain in the context of a competitive model of the labor market as we might hope to able to see the employment effects of such a large change in relative wages. They are more readily explained if labor markets are monopsonistic. The relevant issue is how the marginal cost of female relative to male labor is affected by the Equal Pay Act. One can interpret the requirement not to pay women differently from men as something that is similar to a minimum wage for women which we would expect to reduce the elasticity of the labor supply curve facing firms. In this case, it is possible that changes in the marginal cost of labor are negatively correlated with changes in the wage, thus explaining the result. The bottom line is that the Equal Pay Act seems to have had a substantial impact on the relative wages of women without harming their relative employment.

7.9 Prejudice and Monopsony

The discussion of discrimination so far has focused on the consequences of gender differences in attachment to the labor market. But, there are other potential sources of discrimination. Black (1995) combines the "competitive" model of discrimination of Becker (1971) with its assumption that some agents are racially prejudiced and simply refuse to hire blacks with the Albrecht and Axell (1984) model of monopsony. The consequence of the refusal of some employers to hire blacks is that the effective job offer arrival rate will be lower for blacks, making their labor market less competitive. As one would expect, and as Black (1995) shows, this is to the general disadvantage of blacks and even firms that are not prejudiced end up paying blacks lower wages than whites. Wolpin (1992) and Bowlus and Eckstein (2000) estimate structural models to assess how much of racial differences in economic outcomes can be explained in this way.

There is ample evidence that some racial prejudice remains in studies as diverse as the prices of baseball cards by Nardinelli and Simon (1990) and the performance of English soccer teams by Szymanski (2000). And audit studies where there is some attempt to match job applicants in every characteristic but race (or gender) also suggest that racial discrimination remains in many firms (for a review of the US evidence, see Altonji and Blank 1999; for a recent study, see Bertrand and Mullanaithan 2002). However, these studies are not without their critics (Heckman 1998). At best, they estimate the average level of discrimination. But, in a competitive market, the average level of discrimination is irrelevant. As Heck-

man (1998: 102) puts it, "the impact of market discrimination is not determined by the most discriminatory participants in the market, or even by the average level of discrimination among firms, but rather by the level of discrimination at the firms where ethnic minorities or women actually end up buying." In a frictionless competitive labor market, black workers would have no problem in escaping a wage penalty even if there was only a single non-discriminatory firm (as long as all black workers can be employed there). But, there is an important difference here between competitive labor markets and labor markets with frictions. As soon as there are frictions, the Heckman argument no longer holds: one suffers a real disadvantage as soon as any firm in the market refuses to consider you for employment (a point made by Altonji and Blank 1999). The model of Black (1995) confirms one's intuition on this point. And, any initial disadvantage may be reinforced by social networks (Montgomery 1991b). So, the fact that there are any discriminatory employers in the market is a matter of concern if labor markets are monopsonistic.

7.10 Conclusions

In this chapter we have discussed how an approach to labor markets based on the idea that employers have non-negligible market power can explain certain features of the differences in labor market outcomes for men and women. We have argued that the gender pay gap can be understood as the result of a gender gap in attachment to the labor market and the motivation for job mobility. There is no need to have recourse to differences in productivity between men and women to explain differences in outcomes. Of course, such differences may exist and it is difficult to prove their existence or otherwise. However, the fact that UK equal pay legislation led to a dramatic rise in the relative pay of women without harming their relative employment suggests that the monopsony story may contain some element of truth.

8

Employers and Wages

In a competitive labor market, wages should, after controlling for other relevant characteristics of the worker, only be related to employer and job characteristics to the extent that they affect the non-pecuniary aspects of the job. That is, the only wage variation associated with employers should be compensating wage differentials.

One of the "puzzles" of the observed structure of wages is that wages are correlated with a whole range of employer characteristics (for a survey of this, see Groshen 1991a). One might single out the following:

- industry affiliation (e.g., Krueger and Summers 1988; Gibbons and Katz 1992);
- employer size (e.g., Brown and Medoff 1989; Brown et al. 1990; Oi and Idson 1999);
- profits or profits per worker (e.g., Revenga 1992; Abowd and Lemieux 1993; Blanchflower et al. 1996; Hildreth and Oswald 1997);
- productivity (e.g., Nickell and Wadhwani 1990).

There have been attempts to explain these findings within the competitive model, the main approach taken being to argue that worker quality is observed very badly and that unobserved worker quality may be correlated with employer characteristics (for this type of argument applied to inter-industry wage differentials, see Murphy and Topel 1990). One of these correlations, the employer size–wage effect, has already been investigated in chapter 4 where it was argued that it cannot plausibly be fully explained by a competitive model or a rent-sharing model.

The second section of this chapter shows that the assumption of an upward-sloping supply of labor to the individual employer offers simple and plausible explanations of the empirical regularities listed above. We then turn to compensating wage differentials and discuss the implication of our modeling framework for differences in the non-pecuniary aspects of jobs. We also discuss the likely impact of regulation of the non-wage characteristics of jobs and finish with a discussion of the determinants of hours of work, arguing that working hours are best thought of as simply another non-wage aspect of jobs rather than as a separate topic, that of "labor supply."

8.1 Explaining the Correlations between Employer Characteristics and Wages

This section shows how an upward-sloping supply curve of labor to an individual employer can explain the empirical regularities described in the introduction to this chapter. The employer size–wage effect is, of course, nothing more than another way of saying the supply curve slopes upwards so that empirical regularity is quickly dealt with. But, what about the other correlations?

Consider employers who all face the same supply curve of labor, $N(w)$, but who differ in their revenue function. Denote the revenue function by $Y(N, A)$ where the difference in A is the source of the employer heterogeneity. Assume that $(\partial Y/\partial A) > 0$ so an increase in A is a good shock for the employer.

Obviously each employer will want to choose the level of employment (or, equivalently, the wage) to maximize profits:

$$\pi = Y(N, A) - w(N)N \tag{8.1}$$

leading to the first-order condition

$$\frac{\partial Y(N, A)}{\partial N} = w(N) + \frac{\partial w(N)}{\partial N}N = w(N)[1 + \varepsilon(N)] \tag{8.2}$$

where $w(N)$ is the inverse of the labor supply curve to the firm and $\varepsilon(N)$ is the inverse of the wage elasticity of the labor supply curve facing the firm. The notation has been chosen to make clear that the value of the elasticity might depend on the level of employment chosen.

What is the effect of an increase in A on the wage that the firm pays? The answer depends on how A affects the marginal revenue product of labor $(\partial Y/\partial N)$. If, as seems the most plausible case, an increase in A raises both the average and the marginal revenue product of labor, then an increase in A raises the left-hand side of (8.2) and the optimal wage rises. There is then a positive correlation between the level of profits and the wage paid. A simple diagram makes this clear. Figure 8.1 draws the standard picture of a monopsonistic firm. An increase in the marginal revenue product of labor (MRPL) curve from P_0 to P_1 increases the optimal wage paid from w_0 to w_1. Profitable firms have a high demand for labor and can only get extra labor by paying higher wages. Explaining the positive correlation between the level of profits and the wage paid is embarrassingly simple.

But, what about the correlation between profit per worker and the wage? We are interested in how variations in A affect $(\pi/N) = (Y/N) - w$. Obviously the fact that this depends negatively on the wage makes

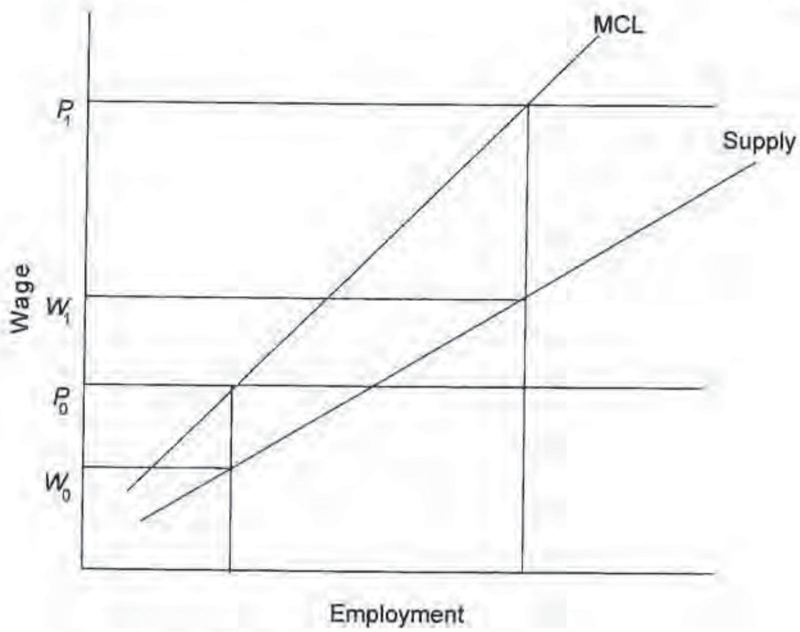

Figure 8.1 Employer heterogeneity and wages.

it harder for the model to explain a positive correlation here. If we denote the elasticity of the revenue function with respect to N by $\alpha(N, A)$, then we have

$$\frac{\pi}{N} = \frac{Y}{N} - w = \frac{1}{\alpha(N,A)}\frac{\partial Y}{\partial N} - w = w\left[\frac{1+\varepsilon(N)}{\alpha(N,A)} - 1\right] \qquad (8.3)$$

(8.3) makes it clear that, if both the revenue function and the labor supply function are iso-elastic, then the model predicts a positive correlation between profits per worker and wages.[1] It is possible to overturn this if the labor supply function is not iso-elastic: in particular, if ε is declining in N as then the right-hand side of (8.3) could conceivably be declining in the wage.[2]

This section has shown how monopsony can explain, rather effortlessly, the correlations between employer characteristics and wages.

[1] One should not claim too much here. If the labor market is competitive with $\varepsilon = 0$, then there is still a positive correlation predicted, something that many studies of the link between profits and wages ignore. The better studies (e.g., Abowd and Lemieux 1993) try to find a variable that affects A and use this as an instrument for profitability.

[2] Perhaps, rather unfortunately, the Burdett and Mortensen (1998) model does have this feature as $(p - w)N$ must be constant for different values of w. This obviously requires that profit per worker $(p - w)$ is decreasing in w. One can reconcile this with (8.3) once one notes that this model has ε as a decreasing function of N.

These correlations are often taken as evidence that the labor market is not competitive and the conclusion is often drawn that wages are above market-clearing levels. But, it is important to realize that we have derived these results in the context of a labor market model in which all workers are being paid below their marginal product. While we have argued that the empirical findings are not consistent with the competitive model, they cannot be used as evidence that wages are above market-clearing levels.

8.2 Monopsony and Compensating Wage Differentials

Non-wage aspects of jobs are important and this section investigates the implications for the structure of wages when jobs differ in their non-pecuniary attributes. The interest in this subject originates in the Wealth of Nations and has assumed considerable importance in discussions about the structure of wages that are based on competitive models. In his survey of the issue in the *Handbook of Labor Economics*, Rosen (1986: 641) wrote that the theory of equalizing differences "can make legitimate claim to be *the* fundamental (long-run) market equilibrium construct in labor economics" (his emphasis). In the competitive model, identically productive workers should receive the same level of utility so that differences in wages should exactly offset differences in the value of non-wage attributes.[3] As Rosen (1986: 641) says, the empirical importance of this result is that it can be used for "making inferences about preferences and technology from observed wage data." The idea is that a competitive labor market should ensure that $U(w, e) = U_0$ where $U(w, e)$ is the utility function, e is the non-wage aspect of the job, and U_0 is the market level of utility. If one controls for labor quality (which will affect U_0) then a regression of w on e should give an estimate of the marginal rate of substitution between e and w. However, compensating wage differentials have not proved of great value in explaining wage variation.[4] The basic problem is that it is hard to find evidence that, other things being equal, more unpleasant jobs are rewarded with higher wages. Often it seems that better-paid jobs have better working conditions.

The most common explanation for these anomalous findings is unobserved worker ability: as Rosen (1986: 671) put it "workers with greater earning capacity would 'spend' some of it on more on-the-job consumption," that is, pleasant working conditions are a normal good (for a

[3] Strictly speaking, this is true only for the marginal worker: if workers differ in the value they attach to non-wage attributes, then there is no single compensating wage differential.

[4] With the possible exception of the returns to education that are sometimes interpreted as a compensating wage differential.

working-out of this intuition, see Hwang et al. 1998). However, it is not clear that this is the root of the problem. For example, Brown (1980) found that, even when panel data are used to control for individual fixed effects (which we might expect to pick up a lot of unobserved ability), it is still hard to find evidence for compensating wage differentials.[5]

This section examines the implications of labor market frictions for the theory of compensating wage differentials and the application of that theory. To keep matters simple, assume all firms have the same productivity p, but that firms differ in the pleasantness of the job offered. Denote the non-pecuniary aspect of the job by e which can be thought to represent effort. Workers are assumed to dislike a high level of e. For the moment, treat e as exogenous: the next section analyzes the case where it is a choice variable for the firm. Assume that workers have a utility function $U(w,e)$ which represents their trade-off between wage income and e. It is natural to think of pleasant work conditions as being a normal good so that, presented with a trade-off between wages and work conditions, workers with higher non-wage income will choose lower wages and better work conditions. This amounts to the condition that

$$[U_{ww}U_e - U_w U_{ew}] > 0$$

In this model the labor supply to the firm will not depend solely on the wage that it offers but on the utility that it offers. Denote the labor supply to a firm that offers utility U by $N(U)$. Also, denote by $w(e,U)$ the wage that needs to be paid by a firm if it wants to offer workers utility U and has working conditions e. Obviously $w_e(e,U) > 0$ and the normality of leisure implies that $w_{eU}(e,U) > 0$ as well. Profits of a firm if it offers utility U will be given by $[p - w(e,U)]N(U)$ and U will be chosen to maximize this.

It is straightforward to show that U must be non-increasing in e so that there is less than full compensation for bad working conditions. The first-order condition for the maximization of profits is given by

$$[p - w(U,e)]N'(U) - w_U(U,e)N(U) = 0 \qquad (8.4)$$

The sign of the response of utility to e is given by the sign of the partial derivative of (8.4) with respect to e (using the second-order condition that we must be at a maximum). So we have

$$\operatorname{sgn}\frac{\partial U}{\partial e} = \operatorname{sgn}[-w_e N' - w_{Ue}N] = \operatorname{sgn}\left[-w_e - \frac{w_{Ue}(p-w)}{w_U}\right] \qquad (8.5)$$

[5] On the other hand, Duncan and Holmlund (1983) do find evidence for compensating wage differentials but that is for Sweden where the wage structure is highly regulated which just goes to prove the dictum that "the only labour markets consistent with the competitive model are regulated labour markets."

where the second equality follows from (8.4). Now $w_e > 0$, and it is simple to check that the normality condition implies that $w_{U_e} > 0$, so that (8.5) implies that high values of e are associated with lower values of U. The implication of this is that workers do not get fully compensated for bad working conditions so that measures of the value attached to working conditions that are based on the assumption of full compensation (and this is the assumption normally made) are likely to be understatements of the true value.

Although the way in which worker utility varies with U is unambiguous, this does not directly tell us about the way that w varies with e. One might expect that a higher value of e is associated with a higher value of w but the higher wage only partially compensates the worker. But, one can construct examples in which a higher e is actually associated with a lower wage.

The model presented so far suggests one reason why estimates of the value of working conditions from earnings functions are so often unsuccessful. The reason is that, in the absence of perfect competition, compensation will not be complete and could even go in the wrong direction. In labor markets with frictions, it is important to remember that even workers of identical ability will receive different levels of utility so that the problem raised here is worse than the problem of unobserved ability that is usually recognized in the competitive model. The existence of frictions can explain why Brown (1980) failed to find evidence of compensating wage differentials even when he controlled for individual fixed effects. Earnings functions alone cannot be relied upon to estimate the value of non-wage job attributes.

This raises the question as to whether there is a potentially better way to estimate the value of working conditions in labor markets with frictions. One appealing way is to consider the estimation of separation functions. As the separation rate depends on worker utility, we might write the separation function in the form $s(U(w,e),x)$ where x represents other relevant factors. The ratio of the coefficients on e and the wage then gives us an estimate of the marginal disutility of bad working conditions. This is the approach taken by Gronberg and Reed (1994) who investigate the impact of four aspects of working conditions (bad working conditions, stooping or kneeling, repetitive work, and heavy lifting). Using the standard hedonic wage equation approach, only bad working conditions have a significant impact in the expected direction. Unfortunately when the impact of these variables is investigated using data on job duration (which will be related to the separation rate), the results are not much better although the disutility of bad working conditions is now much larger as one might expect given the above framework. So, the potential advantage of this approach is not really proved by their study. van Ommeren et al. (1999) use this approach to estimate

TABLE 8.1
The Compensating Differential for Night-Shift Working in the UK LFS

	Coefficient on Log Wage	*Coefficient on Night Work*	*Number of Observations*
Earnings equation	1.00	−0.045 (0.024)	20488
Separations equation	−0.511 (0.051)	0.278 (0.196)	20464
Job-to-job separations equation	−0.525 (0.068)	0.492 (0.241)	19967

Notes.
1. The data come from the Autumn LFS for 1997–2000 inclusive. Night-working is defined as those who report that they work the night shift. Other controls are controls for a quartic in experience, gender, race, education, marital status, the presence of children, and region and month dummies.

the marginal willingness to pay for commuting, and their results are more satisfactory.

Table 8.1 provides an application of this approach to estimating the disutility associated with night-shift working using data from the LFS. Only a small minority of workers (1.5%) report that they work night shifts. The first row shows the results of estimating a standard earnings equation. The coefficient on night-working is negative (although insignificant) suggesting that there is little disutility associated with working nights. The second row uses the Gronberg–Reed approach to estimate a separations equation (as described in section 4.5). The coefficient on night-working is now positive suggesting that, given wages, those working nights are more likely to leave their jobs. However, the estimated coefficient is not significant. The third row restricts attention to job-to-job separations on the grounds that these are more likely to be initiated by the individual and, hence, better reflect their preferences. The coefficient on night-working is larger and now significant at conventional levels. Note that the ratio of the coefficient on night-working to the coefficient on wages gives an estimate of the marginal rate of substitution between earnings and night-working so that these estimates suggest there is a large disutility associated with night-working that is not apparent from a standard earnings function.

8.3 Choice of Working Conditions

So far, working conditions, e, have been assumed exogenous: this section considers the case where e is a choice variable for the firm. A firm will only choose a higher level of e if there is some improved productivity as a result so we assume that productivity is given by

$p(e)$ where $p'(e) > 0$. For simplicity, assume that this function is the same for all firms.

Start by considering what would be the optimal choice of e by workers in perfect competition where they get paid for all the output produced from their effort. They would be interested in maximizing $U(p(e), e)$ which leads to the first-order condition

$$U_w(p(e), e)p'(e) + U_e(p(e), e) = 0 \tag{8.6}$$

Now consider what will happen in a labor market characterized by frictions. The labor supply to firms will depend on the level of utility offered which we continue to denote by $N(U)$. Firms have two instruments to alter the level of utility offered: they can vary the wage and they can vary e. A necessary condition for profit maximization is that, given the level of utility offered by the firm, w and e must maximize $p(e) - w$, that is, firms need to solve the problem

$$\max_{(w,e)} p(e) - w \qquad \text{s.t. } U(w, e) \equiv U \tag{8.7}$$

A necessary condition for profit maximization is that

$$U_w(w, e)p'(e) + U_e(w, e) = 0 \tag{8.8}$$

This is outwardly the same first-order condition that workers would choose in their utopia, (8.6). But, one needs to be careful in drawing the conclusion that the choice of working conditions is efficient. As workers get less than the full value of their contribution to output (w will be less than $p(e)$), effort levels will tend to be higher than they would be in the utopia. But, given the level of utility that workers are going to get, the choice of wages and effort levels is efficient.

To solve for the model, one can then derive e and w as functions of U so that profits can be written as $[p(e(U)) - w(U)]N(U)$. One can then solve for the equilibrium $N(U)$ as all firms must make the same level of profits and the lowest level of utility offered must be the reservation level. Hwang et al. (1998) work out a model of this type more explicitly for a special case where utility is of the form $U(w, e) = w - c(e)$ and where firms differ in the technology of providing e. There is no particular interest in repeating that exercise here except to quote the conclusion of Hwang et al. (1998: 839) that "estimates of workers" marginal willingness to pay derived from the conventional methodology will be biased", a conclusion also reached in the previous section when working conditions were assumed exogenous.

8.4 Mandated Benefits

In a competitive labor market there is little reason to intervene to regulate the conditions under which work is conducted. Yet, in reality, we see many such regulations of the non-wage conditions of work from health and safety regulation, to maximum hours legislation, to parental leave entitlements. All of these types of regulation can be thought of as putting an upper bound on the value of e that is allowed in employment contracts.

Consider the likely impact of such a regulation in a partial equilibrium model of a single firm. First, consider a competitive firm that must pay its workers the market level of utility, U. If it chooses working conditions e, it must pay a wage $w(e, U)$. In the absence of government regulation, e will be chosen to maximize $[p(e) - w(e, U)]$ leading to the first-order condition

$$p'(e) - w_e(e, U) = 0 \tag{8.9}$$

which is the same condition as (8.8) written in a different way.

If the government now intervenes to put a binding upper bound on the value of e, the first effect will be to reduce the profits of the firm without making the workers any better off (which was the presumed intention of the policy). Employers will simply lower wages so that worker utility is still equal to U. In a general equilibrium context, the impact of reduced profits might be to reduce the demand for labor so that the market-clearing level of utility actually falls and workers are made worse off by the well-intentioned intervention (for further discussion, see Summers 1989). But, of course, there are no real grounds for intervention in the first place as the original level of e negotiated between firms and workers is efficient.

What happens in a labor market in which employers have some market power? We have already emphasized that, because workers get less than the full value of their output, there is a tendency for effort to be above the efficient level. This might be thought of as grounds for intervention. But, matters are not that simple because employers may respond to imposed changes in effort by changing wages.[6] Let us assume that the firm faces a supply curve of labor given by $N(U)$. (8.9) will continue to be valid as the first-order condition for the choice of e given U, but the level of utility offered is now also a choice variable for the employer. The first-order condition for the choice of U will be

$$[p(e) - w(e, U)]N'(U) - w_U(e, U)N(U) = 0 \tag{8.10}$$

[6] Of course, they may be prevented from doing so by minimum wages but this is not the situation we are considering here.

Now consider the imposition of a binding upper bound on e. Obviously (8.9) will no longer be valid and the left-hand side will be positive as employers would like to choose a higher value of e. However, (8.10) will still be valid and we can use this first-order condition to work out the impact on U of a change in e. The answer is in the following proposition.

Proposition 8.1

1. *If the restriction on effort is just binding and pleasant working conditions are a normal good, then worker utility must always increase.*
2. *If a binding upper bound is imposed on e, then the sign of the effect on the employer's choice of U is given by*

$$\text{sgn}\left(\frac{\partial U}{\partial e}\right) = \text{sgn}([p'(e) - w_e(e, U)]N'(U) - w_{Ue}(e, U)N(U)) \quad (8.11)$$

Proof. See Appendix 8.

This result, as applied to hours restrictions can be found in de Meza et al. (1998). The first part is, perhaps, surprisingly strong. The intuition for it is that normality implies that, at high levels of effort, a higher increase in the wage is needed to bring about a given increase in worker utility. As higher wages are bad for profits, this acts as a disincentive to raising worker utility. Lowering worker effort can then alter the trade-off between wage costs and employment in such a way as to encourage firms to choose a higher level of utility and employment.

This result does not mean that all intervention will necessarily make workers better off. First, it is a partial equilibrium result and one cannot generalize immediately to a general equilibrium model (although for one way of doing this, see Manning 2001b). Secondly, as one lowers effort there will come a point where worker utility falls. As effort moves further away from the point the employer would choose, $[p'(e) - w_e]$ becomes more positive and this can eventually make the sign of (8.11) positive. Any intervention is likely to be a blunt instrument, laying down common restrictions on e on firms that would otherwise have chosen very different levels of e. Hence, it is likely that the impact is to improve the lot of workers in firms where the constraint just binds, but to worsen it in firms where the constraint is more serious.

Our discussion of working conditions so far has been very abstract. In the next section we study one aspect of working conditions, hours of work, in more detail.

8.5 Hours of Work

One possible interpretation of the e in the previous model is that it represents hours of work and the wage variable represents total earnings. Hours of work are normally modeled differently from other non-wage aspects of jobs but, because, given total labor earnings, hours raise output and reduce worker utility, there is no good reason for this. So, this is the natural place to discuss the implications of our framework for the modeling of labor supply.[7]

In the literature on labor supply, it is conventional to start from a framework in which the worker is faced with a single hourly wage rate and can freely choose hours: what Pencavel (1986: 26) has called the "canonical model". Let us denote the hourly wage rate by w^{h}. The worker will choose e to maximize $U(w^{\mathrm{h}}e, e)$ where e is now hours of work. The first-order condition for utility maximization is given by

$$w^{\mathrm{h}} U_w + U_e = 0 \tag{8.12}$$

It is one of the great unsolved mysteries in labor economics why the canonical model should have received so much attention as there is no particular reason to think that workers would be on their labor supply curve even in a perfectly competitive labor market. To see this, note that perfect competition implies that $w^{\mathrm{h}} = p(e)/e$ but for (8.12) to be consistent with (8.6) requires that $w^{\mathrm{h}} = p'(e)$. These are only mutually consistent if $p(e) = pe$ for some constant p.[8] To give the canonical labor supply model a chance, let us assume that the technology does have this form.

A helpful way of rewriting the profit maximization problem for the firm is to imagine that the choice of the firm is hours of work e and an hourly wage w^{h}. The first-order condition for the maximization problem in (8.7) can then be written as

$$p - w^{\mathrm{h}} + \frac{1}{U_w}\left[w^{\mathrm{h}} U_w + U_e\right] = 0 \tag{8.13}$$

Now we must have $p > w^{\mathrm{h}}$ as otherwise the firm will make no profits. So, (8.13) implies that the term in square brackets must be negative. What this implies is that workers are off their supply curves and being forced to work more hours than they would choose given the single hourly wage

[7] Strictly speaking, we are going to discuss only the intensive margin of labor supply, hours of work. The extensive margin, that between employment, inactivity, and unemployment, is discussed in chapter 9.

[8] A more general result is that the canonical model only holds in a perfectly competitive equilibrium if the production function can be written as $f(Ne)$, that is, output is a function solely of total hours worked. This point was first made by Lewis (1969): see Manning (2001c) for further discussion.

rate. One implication of this is that it is not worker preferences alone that will determine hours of work: employer characteristics will also be important. The parallel here is with the earnings equation literature where, under perfect competition, wages will be unrelated to employer characteristics but, as we have seen, that is not the case when employers have some market power. A similar result is given in (8.13): if we have perfect competition then $p = w^h$ and worker preferences alone will be important in determining hours of work, but if employers have some market power this is no longer the case.

There are a number of ways in which workers might be forced off their classical labor supply curves. One approach is simply to dictate a wage–hours package, another to offer a non-linear relationship between hours and wage. In reality, both of these strategies are used.

Table 8.2 presents some evidence on the extent of over-employment. In the BHPS, individuals are asked whether, at their hourly wage, they want to work more hours, fewer hours, or continue as they are. 33% of workers want to work fewer hours as opposed to 9% who want to work more. Men are more likely than women to want fewer hours. This is supportive of the view that there is a tendency towards excessive hours (from the worker's point of view). However, it is not consistent with a literal interpretation of the result in (8.13) that *all* workers should be over-employed. One needs to explain why some workers may be under-employed and others content with their lot.

TABLE 8.2
Desired Hours: BHPS

	Work Fewer	*Continue the Same*	*Work More*	*Sample Size*
Percentages				
All	32.1	60.0	8.9	20630
Men	34.8	57.0	8.2	9661
Women	29.7	60.8	9.4	10969
Average hours				
All	38.2	32.9	26.4	20630
Men	41.0	38.8	35.5	9661
Women	35.3	28.0	19.5	10969

Notes.
1. The data in this table come from the BHPS for 1991–98 and tabulate the responses for those in employment to the question "Thinking about the hours you work, assuming that you would be paid the same amount per hour, would you prefer to work fewer hours, more hours or continue with the same hours." The average hours are the number of hours normally worked per week.

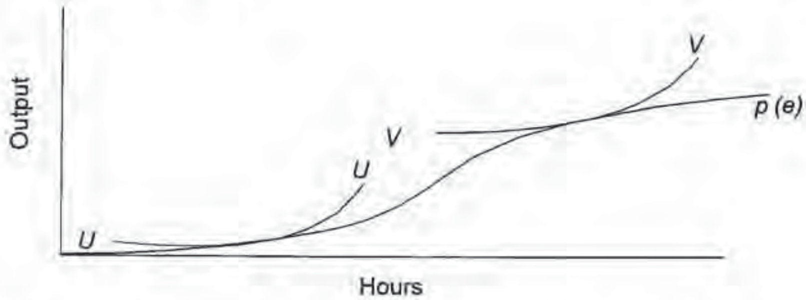

Figure 8.2 Under- and over-employment in a competitive labor market.

First, there is a tendency in many jobs for hours of work to be equalized across workers because of the technological importance of having workers "on-site" at the same time. The standard working day is the best example of this. With heterogeneity in preferences, this will inevitably result in some workers being under-employed even if, on average, workers are over-employed. Secondly, the result that individuals are at their desired hours conditional on their average hourly wage is only true in a competitive model if the production function can be written as linear in hours. If it is not, then individuals will tend to be over- or under-employed. Figure 8.2 clarifies this point. Suppose the relationship between output produced and hours is given by the $p(e)$ function. The justification for the shape is that there are "set-up" costs associated with the job so that output is initially a convex function of hours but that exhaustion means that there are eventually diminishing returns to hours. In a competitive labor market, the chosen point will be a point of tangency between an indifference curve and this production function. Workers will feel over-employed if the average wage exceeds the marginal wage and under-employed if the opposite is the case. So, someone with preferences *UU* who is working a small number of hours would feel over-employed while someone with preferences *VV* who is working a lot of hours would feel under-employed. Given the shape of the production function, there is a tendency for those employed at high hours to feel under-employed and those employed at low hours to feel under-employed. As the second half of table 8.2 shows, this is not what we see in the data. This makes no attempt to control for characteristics but the study of Stewart and Swaffield (1997) comes to the same conclusion. It is those who work long hours who are more likely to feel over-employed. So, a competitive model with a more flexible specification of the production function does not seem able to explain the data.

Table 8.2 also shows that a majority of workers are happy with the hours they are working. Stewart and Swaffield (1997) interpret these

individuals as working the exact number of hours they desire. However, it may be that these responses reflect an acceptance of the labor market realities, that is, satisficing rather than optimizing. For example, individuals are also asked how satisfied they are with the pay in their job: 60% of respondents say that they are at least satisfied with their level of pay. Yet, I suspect that most of these would happily accept a pay rise.

This discussion raises the issue of how an empirical investigation of the determinants of hours of work should be conducted. There is a huge amount of existing research (summarized in Blundell and MaCurdy 1999), the vast bulk of which starts from the canonical model and assumes that workers have a free choice of hours of work and that the average hourly wage paid to them is also their marginal hourly wage (abstracting from taxes). It is a pity that quite so much attention should have been lavished on the canonical model for, as we have already discussed, there is little reason to think that a perfectly competitive labor market would have workers on their classical labor supply curves.

It is perhaps best to think of hours of work as just another non-wage attribute of jobs and, if one believes in a competitive labor market, to use the compensating wage differentials literature to estimate the marginal rate of substitution between income and hours. Suppose that individuals have preferences $U(w, e, x_f)$ where w is total earnings, e is hours of work and x_f is other characteristics affecting preferences. In a competitive labor market, workers of a given quality, x_q, will get a level of utility $U_0(x_q)$ so that the trade-off between income and hours will be given by $U(w, e, x_f) = U_0(x_q)$. A regression of log earnings on a suitable function of hours and other characteristics could then give an estimate of the marginal rate of substitution between income and hours. If w is a linear function of hours, then this reduces to the canonical case.

However, we discussed above why this conventional approach to estimating marginal willingness-to-pay is unlikely to give the correct answer if employers have some market power. In the current context, if leisure is a normal good, then those in high-paying jobs are, other things being equal, likely to be in jobs with lower hours. An earnings function will then tend to underestimate the disutility of work. As an alternative, we could use the Gronberg–Reed methodology and relate separations to earnings and hours of work.

Table 8.3 shows what happens when we do this using data from the US PSID and the UK LFS and BHPS. First, consider the results using the PSID in table 8.3a. The first row is an estimate of an earnings function for men where the dependent variable is the log of weekly earnings and the log of hours is included on the right-hand side. The estimate of 1.051 can be thought of as a crude estimate of the marginal rate of substitution

TABLE 8.3a
Labor Supply: PSID

Sample	*Equation*	*Coefficient on Log (Weekly Earnings)*	*Coefficient on Log (Hours)*	*Coefficient on Log (Hours) (Hours < 30)*	*Coefficient on Log (Hours) (Hours > 30)*	*Number of Observations*
Men	Earnings function	1.000	1.051 (0.018)			31471
Men	Separations function	−0.838 (0.062)	0.604 (0.108)			25343
Men	Job-to-job separations	−0.838 (0.062)	0.827 (0.096)			22315
Men	Earnings function	1.000		1.208 (0.036)	0.954 (0.016)	31471
Men	Separations function	−0.855 (0.060)		0.806 (0.096)	0.429 (0.172)	25343
Men	Job-to-job separations	−0.893 (0.043)		0.993 (0.176)	0.720 (0.151)	22315
Women	Earnings function	1.000	1.146 (0.009)			29295
Women	Separations function	−0.838 (0.062)	0.610 (0.081)			24319
Women	Job-to-job separations	−0.838 (0.062)	0.821 (0.138)			19864
Women	Earnings function	1.000		1.102 (0.017)	0.954 (0.016)	29295
Women	Separations function	−0.972 (0.052)		0.487 (0.091)	0.914 (0.108)	24319
Men	Job-to-job separations	−0.914 (0.066)		0.564 (0.148)	1.258 (0.156)	19864

Notes.
1. The other controls included are a quartic in experience and job tenure, education, race, marital status, the presence of children, year, and state dummies.
2. For details of how the separation equations are estimated, see section 4.5.

TABLE 8.3b
Labor Supply: LFS

Sample	*Equation*	*Coefficient on Log (Weekly Earnings)*	*Coefficient on Log (Hours)*	*Coefficient on Log (Hours) (Hours < 30)*	*Coefficient on Log (Hours) (Hours > 30)*	*Number of Observations*
Men	Earnings function	1.000	0.831 (0.009)			35731
Men	Separations function	−0.464 (0.042)	−0.012 (0.076)			30383
Men	Job-to-job separations	−0.493 (0.053)	0.152 (0.109)			29770
Men	Earnings function	1.000		1.254 (0.015)	0.497 (0.013)	35731
Men	Separations function	−0.449 (0.042)		−0.229 (0.101)	0.302 (0.126)	30383
Men	Job-to-job separations	−0.479 (0.054)		−0.082 (0.165)	0.359 (0.158)	29770
Women	Earnings function	1.000	1.048 (0.004)			37218
Women	Separations function	−0.526 (0.040)	0.049 (0.055)			31809
Women	Job-to-job separations	−0.432 (0.058)	0.230 (0.083)			30867
Women	Earnings function	1.000		1.100 (0.006)	0.841 (0.016)	37218
Women	Separations function	−0.523 (0.040)		0.018 (0.062)	0.203 (0.157)	31809
Men	Job-to-job separations	−0.437 (0.058)		0.276 (0.101)	0.871 (0.214)	30867

Notes.
1. The other controls included are a quartic in experience and job tenure, education, race, marital status, the presence of children, year, and region dummies.
2. For details of how the separation equations are estimated, see section 4.5.

TABLE 8.3c
Labor Supply: BHPS

Sample	*Equation*	*Coefficient on Log (Weekly Earnings)*	*Coefficient on Log (Hours)*	*Coefficient on Log (Hours) (Hours < 30)*	*Coefficient on Log (Hours) (Hours > 30)*	*Number of Observations*
Men	Earnings function	1.000	0.761 (0.037)			15758
Men	Separations function	−0.765 (0.078)	0.336 (0.154)			9189
Men	Job-to-job separations	−0.743 (0.107)	0.875 (0.257)			8513
Men	Earnings function	1.000		1.224 (0.062)	0.314 (0.048)	15758
Men	Separations function	−0.689 (0.078)		−0.349 (0.202)	1.070 (0.209)	9189
Men	Job-to-job separations	−0.663 (0.107)		−0.230 (0.351)	1.466 (0.265)	8513
Women	Earnings function	1.000	1.111 (0.014)			16582
Women	Separations function	−0.761 (0.066)	0.329 (0.089)			10337
Women	Job-to-job separations	−0.704 (0.100)	0.449 (0.144)			9204
Women	Earnings function	1.000		1.164 (0.019)	0.860 (0.069)	16582
Women	Separations function	−0.752 (0.066)		0.253 (0.101)	0.689 (0.243)	10337
Men	Job-to-job separations	−0.693 (0.100)		0.360 (0.164)	0.802 (0.367)	9204

Notes.

1. Data are from the BHPS 1991–98. The other controls included are a quartic in experience and job tenure, education, race, marital status, the presence of children, year, and region dummies.
2. For details of how the separation equations are estimated, see section 4.5.

between log hours and log income in the utility function.[9] However, when we estimate a separations equation using as dependent variable an indicator taking the value 1 if the individual left their job and 0 if they did not, then, as we would expect, separations are less likely if earnings are high and hours are low. However, the estimated marginal rate of substitution is lower than that implied by the earnings function at 0.721 (=0.604/0.838). However, if attention is restricted to job-to-job separations (on the grounds that these are more likely to be voluntary), the estimated marginal rate of substitution rises to 0.987. For women, in the PSID, the results are qualitatively similar. These results do not suggest that the standard earnings function approach understates the marginal rate of substitution.

For the United Kingdom, the results are rather different. Table 8.3b reports the results from the LFS and table 8.3c those from the BHPS. For men in the LFS, the coefficient on log hours in the separations equation is lower than in the PSID and often insignificant. However, this seems to be the result of a non-linearity. If a spline is estimated using 30 hours as the break point, the results suggest that higher hours do lead to higher separations when hours are high but not when hours are low. At low levels of hours, an increase in hours makes British men less likely to quit, evidence perhaps that men dislike part-time work as it is overwhelmingly "female" and undermines the perception of their "masculinity." Table 8.3c confirms the existence of this effect for the BHPS. However, as shown in some rows of table 8.3a, US men show no evidence of this effect. For women, this effect is much less marked. Separation rates seem to be increasing in hours whatever the level of hours worked. But, it remains the case that the approach to estimating the marginal rate of substitution based on separations equations does not always lead to a higher estimate of the disutility of work so here, as in the previous analysis of night-working, the Gronberg–Reed approach does not perform as one might expect.

This section has been excessively cursory to economize on space but it does suggest an alternative approach to estimating "labor supply curves" from that which is most commonly used in the literature. A more convincing analysis should pay attention to the impact of tax and welfare systems on the utility of individuals and use more general specifications of the utility function.

8.6 Conclusion

This chapter has shown how a model of the labor market in which employers have some market power can explain the correlations between

[9] We do not experiment here with alternative functional forms of the utility function or with adequate treatment of the tax system, the sophisticated treatment of which is the hallmark of much empirical work on labor supply (see Blundell and MaCurdy 1999).

employer characteristics and wages which have often been regarded as puzzles that need to be explained away. It has also shown how a monopsonistic perspective has important implications for how one estimates the value attached to working conditions, for the determination of hours, and for the likely impact of mandated benefits. It has argued that an alternative approach to labor supply is to regard hours of work as a "disamenity" and to estimate the marginal rate of substitution between hours and wages using the methodology proposed for other compensating wage differentials.

Appendix 8

Proof of Proposition 8.1

Write the left-hand side of (8.10) as $\pi_U(U, e)$ where the notation captures the idea that this is the first-order condition for profits with respect to utility. This must be equal to zero even with the imposition of a binding regulation on e. Hence, the way in which utility must vary with a change in e is given by

$$\pi_{UU}\frac{\partial U}{\partial e} + \pi_{Ue} = 0 \qquad (8.14)$$

π_{UU} must be negative by the second-order condition for (8.10) to give us the profit-maximizing level of U. Hence, we have that $\text{sgn}(\partial U/\partial e) = \text{sgn}(\pi_{Ue})$. Differentiating (8.10) then gives us (8.11) which proves part (2). Now, if the constraint is just binding, (8.9) must be satisfied which, using (8.11) means that $\text{sgn}(\partial U/\partial e) = -\text{sgn}(w_{Ue})$ which is negative if pleasant working conditions are a normal good. This proves part (1).

Part Three

LABOR DEMAND AND SUPPLY

9

Unemployment, Inactivity, and Labor Supply

This chapter discusses the determinants of the level and structure of non-employment. In a frictionless, perfectly competitive labor market, workers are out of work whenever the utility they could obtain in the market is below the utility obtainable when out of it. In most discussions, however, allowance is made for some frictional component to unemployment.[1] Perhaps the most celebrated statement of this is Friedman (1968: 8)

> the natural rate of unemployment is the level ... that would be ground out by the Walrasian system of general equilibrium equations, provided there is embedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility and so on.

Consequently, there is a large literature that uses a search approach to analyze unemployment (for surveys, see Mortensen 1986; Mortensen and Pissarides 1999). The consequence of this is that, as mentioned at the end of chapter 2, thinking about unemployment using search theory is much more familiar than thinking about the wage distribution using these tools. So, there is even less that is "new" in this chapter than in the preceding ones.

Using the notation of the basic Burdett and Mortensen (1998) model introduced in section 2.4, the non-employment rate of a worker with reservation wage r can be written as

$$u = \frac{\delta_u}{\delta_u + \lambda(1 - F(r))} \tag{9.1}$$

where δ_u is the rate of job destruction, λ is the rate at which job offers arrive when unemployed, $F(w)$ is the wage offer distribution, and r is the reservation wage.[2] In the basic model of section 2.4, δ_u and λ were

[1] As discussed in chapter 1, it is unclear how one can reconcile a belief in frictions in labor markets with a belief in a perfectly competitive method of wage determination.

[2] In the version of the model used in section 2.4, $F(r) = 0$ as all workers were assumed to have the same reservation wage and there was no point in any employer offering a wage lower than that. The specification given here is more appropriate when there is heterogeneity in reservation wages as in the version of the model discussed in section 3.5.

assumed exogenous, and the reservation wage was equal to the flow of utility obtainable when non-employed because it was assumed that on- and off-the-job searches were equally effective in generating job offers. A more adequate discussion of non-employment requires more attention to all of these parameters. The transition rates can obviously be influenced by the actions of both workers and employers. The arrival rate of job offers will be affected both by the level of search intensity of a non-employed worker and by the recruitment intensity of employers. And, the separation rate will be affected by both the quit decisions of workers and the lay-off decisions of employers. This chapter considers the workers' actions that are likely to affect the transition rates while the next chapter considers the actions of employers.

The plan is as follows. The first section of this chapter allows individuals to choose the intensity of job search and discusses the implications of this for the non-employment rate and the determinants of the reservation wage. Evidence is presented to suggest that off-the-job search is more effective than on-the-job search. The second section discusses the difference between unemployment and inactivity, a distinction commonly made in labor market statistics. As these two labor market states are distinguished primarily using a measure of search intensity (the unemployed being those who have looked for work in the recent past and the inactive those who have not), a model of endogenous search intensity can provide a meaningful theory of the difference between inactivity and unemployment. This approach is described in the second section where there is an application to the discouraged worker effect.

The third section discusses the job search intensity of the employed. This does not directly affect the non-employment rate in (9.1) but this seems the natural place to discuss it. Theory predicts that on-the-job search intensity should be negatively related to wages and empirical evidence strongly confirms this. The fourth section discusses the quit behavior of workers showing how a model with a stochastic reservation wage can readily explain the fact that quits to non-employment are negatively related to the wage.

Another aim of this chapter is to improve the understanding of the nature of unemployment in an oligopsonistic labor market. In the basic models of monopsony used so far, all firms are trying to recruit workers all the time and a match between a worker and employer only ever fails to be consummated because the wage offered is below the reservation wage of the worker. This does not seem to give an adequate description of the difficulties that many of the unemployed have in finding work. So, the fifth and sixth sections of this chapter shows how one can marry conceptions of involuntary unemployment, efficiency wages and monopsony. It turns out that this is not problematic.

9.1 Endogenizing Job Search Activity

An individual will determine the level of job search by equating marginal benefits and costs. Denote the cost to a non-employed individual of generating job offers at a rate λ by $c_u(\lambda, z)$ where z represents other relevant factors. Some of these other factors may be individual-specific (e.g., the appeal of the individual to employers) while others may be beyond their control (e.g., the state of the aggregate labor market). Similarly define $c_e(\lambda, z)$ to be the cost to an employed individual of generating job offers at a rate λ for employed workers. The cost functions for employed and non-employed individuals may differ for a number of reasons: non-employed workers have more time (an advantage) but less money (a disadvantage), or employers may discriminate against non-employed job applicants (for some UK evidence in support of this view, see Manning, 2000). The relative importance of on-the-job and off-the-job searches turns out to be of some importance and is discussed further below.

Denote by λ_u the arrival rate of job offers for an unemployed worker. A non-employed worker will choose λ_u to maximize the value of being unemployed, V^u, which will be given by

$$\delta_r V^u = \max_{\lambda_u} b + \lambda_u \int_r [V(x) - V^u] dF(x) - c_u(\lambda_u, z) \qquad (9.2)$$

where $F(x)$ is the wage offer distribution, r is the reservation wage, and $V(x)$ is the value of a job that pays wage x. The first-order condition for this maximization problem can be written as

$$\frac{\partial c_u(\lambda_u, z)}{\partial \lambda_u} = \int_r [V(x) - V^u] dF(x) \qquad (9.3)$$

The right-hand side of (9.3) is the expected gain from taking a job above the reservation wage so (9.3) says that the marginal cost of an extra job offer should be equal to the marginal expected benefit from one. The reservation wage appears in (9.3) although one should remember that this is endogenous. Some comparative statics are simple: if the level of utility obtainable when out of employment, b, rises (perhaps because of a rise in welfare benefits), then the reservation wage will rise and (9.3) says that the search intensity of the unemployed will fall (as has been claimed by Barron and Mellow 1979).[3]

[3] It is worth noting that a literature has also grown up surrounding the lack of robustness of this conclusion emphasizing how, in a world where the unemployed are liquidity-constrained (Hamermesh 1982; Ben-Horim and Zuckerman 1987) or leisure is locally inferior (van den Berg 1990), increased benefits may lead the unemployed to search more intensively. Blau and Robins (1990), Wadsworth (1991) and Schmitt and Wadsworth (1993) find evidence that benefit entitlement increases search effort. Of course, this may be the conse-

Now, consider the choice of search intensity by the employed. The value function for an employed worker can be written as

$$\delta_r V(w) = \max_{\lambda_e} w + \lambda_e \int_w [V(x) - V(w)]dF(x) - \delta_u(V(w) - V^u) - c_e(\lambda_e, z) \tag{9.4}$$

leading to the first-order condition

$$\frac{\partial c_e(\lambda_e, z)}{\partial \lambda_e} = \int_w [V(x) - V(w)]dF(x) \tag{9.5}$$

Denote the solution to this by $\lambda_e(w)$ as the solution will depend on the wage (this is discussed in section 9.3). The following expression for the reservation wage can be derived.

Proposition 9.1. *The reservation wage, r, is given by the solution to*

$$r + (\lambda_e(r) - \lambda_u) \int_r \frac{[1 - F(x)]dx}{\delta + \lambda_e(x)[1 - F(x)]} - c_e(\lambda_e(r), z) = b - c_u(\lambda_u, z) \tag{9.6}$$

Proof. See Appendix 9.

There are a number of uses to which the reservation wage of equation (9.6) can be put, but also certain potential uses of it that are a bit misleading as it is based on a partial equilibrium model in which the wage distribution is treated as fixed. Consider, for example, the argument that the payment of welfare benefits when out of work reduces work incentives and leads to a lower employment rate. At one level, (9.6) supports this prediction as an increase in unemployment insurance payments for an individual will raise the reservation wage and reduce search intensity, both effects tending to increase the non-employment rate. But, caution is needed before generalizing this conclusion to an economy-wide increase in welfare benefits. The problem is that this conclusion is based on the assumption that the distribution of wages is fixed. However, a general increase in welfare benefits is likely to affect the distribution of wages (for an example of this, see section 2.4) in which case it is not necessarily true that the non-employment rate rises. An interesting application of this idea can be found in van Vuuren et al.

---

quence of the fact that benefit receipt is often made conditional on a certain level of search intensity. It is simple to show in this case that some workers who would otherwise search less than the threshold will search at the threshold intensity and so can only increase their search intensity, while those who were initially above the threshold will reduce their intensity though not so much as to fall below the threshold.

(2000) who consider what happens when there is dispersion in unemployment benefits.

Some other features of (9.6) also deserve discussion. One special case is where the functions $c_u(\lambda, z)$ and $c_e(\lambda, z)$ are identical: in this case comparison of (9.3) and (9.5) (and using the fact that $V(r) = V^u$) shows that we will have $\lambda_e(r) = \lambda_u$ and (9.6) then implies that $r = b$. So, if on- and off-the-job searches are equally effective, the reservation wage will equal the disutility of leisure. Note that this does not mean that all employed workers will search at the same intensity as the non-employed, just that those at the reservation wage will.

Whether on- or off-the-job search is more effective is a question that economists have worried about and several substantive issues depend on the answer to this question (e.g., see the discussion about the impact of trade unions on non-union wages in chapter 12). The existing literature on the subject is not particularly satisfactory. For example, Holzer (1987) finds, using the NLSY, that the unemployed use more search methods than employed job searchers, spend more time on it, get and accept more job offers. He concludes that off-the-job is more effective than on-the-job search. In contrast, Blau and Robins (1990) find that employed job seekers are slightly more successful than unemployed job seekers in getting offered jobs. However, as only 10% of employment spells are associated with job search, and we are interested in the average rate at which employed workers change jobs, one might reasonably argue that their study suggests that off-the-job search is more effective. Both these papers have the problem that all employed job seekers are lumped together whereas (9.6) shows that one should compare unemployed job seekers with those employed at the reservation wage.[4]

Here we take a different approach to the problem using the reservation wage equation (9.6). Suppose we consider a variable, call it labor quality, q, which raises the wage offer distribution in the sense of first-order stochastic dominance. Represent the wage offer distribution by $F(w, q)$: the assumption of first-order stochastic dominance implies that $F_q(w, q) < 0$. By differentiating (9.6) it is simple to show that a rise in q will raise (reduce) the reservation wage if off-the-job search is more (less) effective than on-the-job search. However, average wages will be increasing in labor quality in both cases. This suggests testing the relative effectiveness of on-the-job and off-the-job search by investigating the impact of variables that raise labor quality on the reservation wage.

The UK British Household Panel Survey (BHPS) asks those who are looking for work but do not currently have a job about their reservation

[4] If those employed at higher wages search for another job less intensively, this may not be because they are at a disadvantage in job search but because the perceived returns on job search are smaller. This hypothesis is tested below.

wage. We can use this information to estimate a reservation wage equation. This equation is like a standard human capital earnings function but with the addition of variables that we think might affect b, the value of leisure. We include investment income and welfare benefit income in the previous month (converted to an hourly basis by dividing by 170). We also include a dummy variable for whether the household is in receipt of family credit, a benefit that is designed to also be received by those in work. The first column of table 9.1 presents a standard earnings function although we include the investment income and benefit variables for comparison. The findings are familiar: earnings increase with education (the omitted education category is those with a college degree), and are a concave function of experience (although flatter for women), there is a pay premium for married men, and a pay penalty for women with children. Earnings are strongly positively related to investment income (which probably reflects high past earnings) and negatively with benefit

TABLE 9.1
The Determinants of the Reservation Wage

Dependent Variable	*Log Hourly Wage*	*Log Reservation Wage*
"A" levels	−0.373 (0.010)	−0.253 (0.029)
GCSEs	−0.546 (0.010)	−0.351 (0.028)
No qualifications	−0.712 (0.012)	−0.432 (0.029)
Female	0.114 (0.047)	0.031 (0.066)
Experience	0.046 (0.001)	0.028 (0.002)
Experience squared	−0.00085 (0.00003)	−0.00049 (0.00003)
Experience × female	−0.021 (0.002)	−0.013 (0.003)
Experience squared × female	0.00032 (0.00004)	0.00024 (0.000050)
Married	0.161 (0.013)	0.141 (0.023)
Married × female	−0.139 (0.017)	−0.142 (0.030)
Children	0.027 (0.011)	−0.029 (0.024)
Children × female	−0.125 (0.015)	0.046 (0.032)
White	0.104 (0.032)	−0.049 (0.045)
White × female	−0.052 (0.045)	−0.046 (0.060)
Health problems	−0.076 (0.013)	−0.026 (0.015)
Household investment income	0.023 (0.002)	0.015 (0.007)
Household benefit income	−0.020 (0.002)	0.004 (0.002)
Family credit receipt	−0.225 (0.026)	−0.092 (0.039)
Number of observations	15393	3634
R^2	0.43	0.23

Notes.
1. The sample is drawn from the BHPS 1991–98. Regional and year dummies are also included. The omitted education category is a college degree.

receipt (which reflects the rules of benefit entitlement). The second column now estimates a reservation wage equation. What is striking is the similarity of the coefficients although, as we might expect, benefit income now has a positive impact on the reservation wage.[5] The impact of the "quality" variables, education and experience, suggests that off-the-job search is more effective than on-the-job search as we would otherwise expect measures of labor quality to have opposite impacts on the two equations. This is not to say that there are no parts of the labor market where the reverse may be true (it is probably harder to get a job as a university lecturer if one is unemployed): just that, on average, it is easier to get a job when unemployed.

9.2 Unemployment and Inactivity

The discussion so far has not made a distinction between unemployment and inactivity: there is only a distinction between employment and non-employment. Labor market statistics often divide the non-employed into two groups, the unemployed and the inactive, using a definition provided by the International Labour Organization. To be classified as unemployed rather than inactive, one must not have worked at all in the reference week or be temporarily away from a job and:

either have looked for work in the past four weeks
and be available to start work within two weeks
or be waiting to start a job within two weeks

In practice the first definition is the most important: only 2% of the unemployed in the United Kingdom fail the first test but pass the second. The purpose of the definition is to distinguish those who want a job but have not got one (a cause of public policy concern) from those who do not want a job. In recent years, this clear-cut distinction has become somewhat blurred, for example, there has been a striking rise in the inactivity rates of less-educated prime-age males in the both the United States (Juhn et al. 1991) and the United Kingdom (Nickell and Bell 1995, 1996). Many of these are inactive because of reported health problems (e.g., see Autor and Duggan 2001) yet it is hard to see what epidemic has been sweeping these countries making millions of men too sick to work. If their labor market was in a healthier state, many of these would almost certainly be in employment.

Models of labor supply based on competitive models of the labor market find it hard, if not impossible, to have a meaningful distinction

[5] Although its impact is small; at the average level of benefit income (£1.72 per hour), the elasticity of the reservation wage with respect to benefits is 0.015.

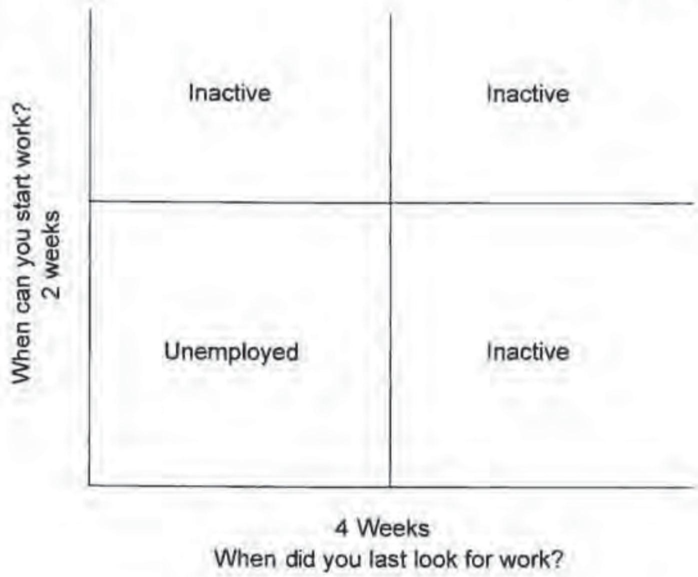

Figure 9.1 The distinction between inactivity and unemployment.

between unemployment and inactivity. The problem is that, in a frictionless labor market, workers can always get a job immediately if they want one. The process of looking for work does not really exist and anyone who is not currently working is assumed not to want a job in spite of any protestations to the contrary. There are some labor supply researchers who feel uneasy with this implication of their approach (e.g., see Blundell et al. 1987), but no entirely satisfactory way of dealing with the problem has been developed.

Search models are ideally suited to modeling the distinction between inactivity and unemployment as the process of "looking for work" is the core of the model. Labor force surveys typically ask the question "have you looked for work in the past four weeks?" and "are you available to start work within two weeks?" but imagine that the questions were "when did you last look for work?" And "when is the earliest you could start work?". One could plot the answers to these questions in figure 9.1. The unemployed will be those with answers in the bottom left-hand box: all the others will be classed as inactive. In practice, the search intensity requirement seems to be more important than the job availability requirement[6] so that modeling the distinction between unem-

[6] Those who are searching for work but not available to start a job within 2 weeks tend to be specific groups like those in full-time education or caring for children who have a good idea of when they will want a job in the future.

ployment and inactivity is primarily a question of modeling search intensity. This is the approach taken in Burdett and Mortensen (1978), Burdett et al. (1984), and Blundell et al. (1998) who all model a search–no search decision.

Figure 9.1 should make it clear that the discrete distinction between the unemployed and the inactive is an arbitrary classification placed on what is really a continuous underlying measure of attachment to the labor market. There is a literature, starting with Flinn and Heckman (1983), that asks whether unemployment and inactivity are distinct labor market states and tries to answer this question by examining labor market transition rates. A glance at figure 9.1 should make it clear that this is not a well-posed question: on average, the inactive are less likely to enter employment than the unemployed but, at the margin, the transition rates are likely to be identical. This view is confirmed by the studies of Jones and Riddell (1999) and Gregg and Wadsworth (1996) who find that some of the inactive are more attached to the labor market than others. It is not the mystery sometimes claimed that there are direct moves from inactivity into employment as some of the inactive have low but non-zero job search activity. One should not think of the inactive as having no attachment to the labor market, just a low attachment.

Once one recognizes that the conventional classification of the non-employed into the inactive and the unemployed is an arbitrary one, one begins to wonder whether it is a helpful distinction or the best way of classifying the non-employed into different categories. Jones and Riddell (1999) divide the inactive as conventionally measured into two groups, which they call the marginally attached and the inactive. But, one could go further in subdividing the inactive. The logical end-point of this subdivision is to say that one would like to know the full distribution of λ_u. Alternatively, one could start from the premise that, in the interest of having published statistics that are easily understood, we want to divide the non-employed into a small number (say two) of groups (which we will call the unemployed and the inactive), and ask whether the current way of doing this classification is the best, given the information available on attachment to the labor market.

One issue that has interested labor economists in the past is whether a worsening in the general state of the labor market tends to increase labor force participation (the added-worker effect) or reduce it (the discouraged worker effect). A worsening in the labor market can be thought of as saying that, for a given level of search intensity by the individual, job offers arrive at a slower rate than before so that $c_z < 0$. But, from (9.3), one can see that a lower level of z leads to a lower (higher) level of λ_u as $(\partial^2 c_u/\partial z \partial \lambda_u) > (<) 0$. The sign of this cross-partial derivative is unclear a priori.

If job offers arrive more slowly than before this may either increase or reduce the marginal value of a given level of search intensity. It is possible that in bad times job offers are so hard to find that it is simply not worth trying: this is one version of the discouraged-worker effect. Or, it may be the case that, in good times, job offers arrive at a fast rate with very little effort and extra effort brings little extra reward: this is a version of the added-worker effect. The main interest in this type of exercise is whether we should think of search intensity as pro- or counter-cyclical or whether the discouraged- or added-worker effect dominates.

Figure 9.2 presents some evidence on this. Figure 9.2a plots the inactivity rate for men aged 45–60 against a measure of the state of the local labor market, the employment/population ratio for prime-aged men (those aged 25–44) for US states (excluding Alaska) for the period 1998–2000 inclusive. There is a negative relationship between the two suggesting that when the labor market is worse in a state, more older men move out of the labor force into inactivity. The regression line is also marked on the figure: the slope coefficient is −1.04 with a standard error of 0.15. Figure 9.2b plots the same relationship for UK counties over the period 1998–99 inclusive. Again, there is a negative relationship: this time the slope coefficient is −1.17 with a standard error of 0.09. This evidence suggests that when labor markets worsen, some older men reduce their search intensity so much that they exit the labor force and are no longer classified as unemployed.

But, one might wonder whether there is any evidence that search intensity also falls among those who remain in the labor force. Gregg and Wadsworth (1996), using British data from the 1980s and early 1990s suggest, rather tentatively, that, among job searchers, the number of job search methods used increases as the labor market worsens. Some further evidence on this point from the US CPS and the UK LFS is presented in table 9.2. It is obviously hard to measure search intensity: the best measure commonly available is the number of search methods used. For those who are inactive this will (except for the small group who are unavailable for work) be equal to zero, while for those who are unemployed it must be greater than zero. Table 9.2 investigates the impact of the state of the local labor market on the average number of search methods used by those in the labor force. As in figure 9.2, the state of the local labor market is measured by the prime-age (25–54) male employment/population ratio for the region in which the individual lives. The results are very similar for the United Kingdom and the United States. Most of the specifications (which include varying combinations of personal, regional, and time controls) suggest a significant positive effect of the employment/population ratio on the number of search methods used. This is consistent with the earlier conclusion that a good labor

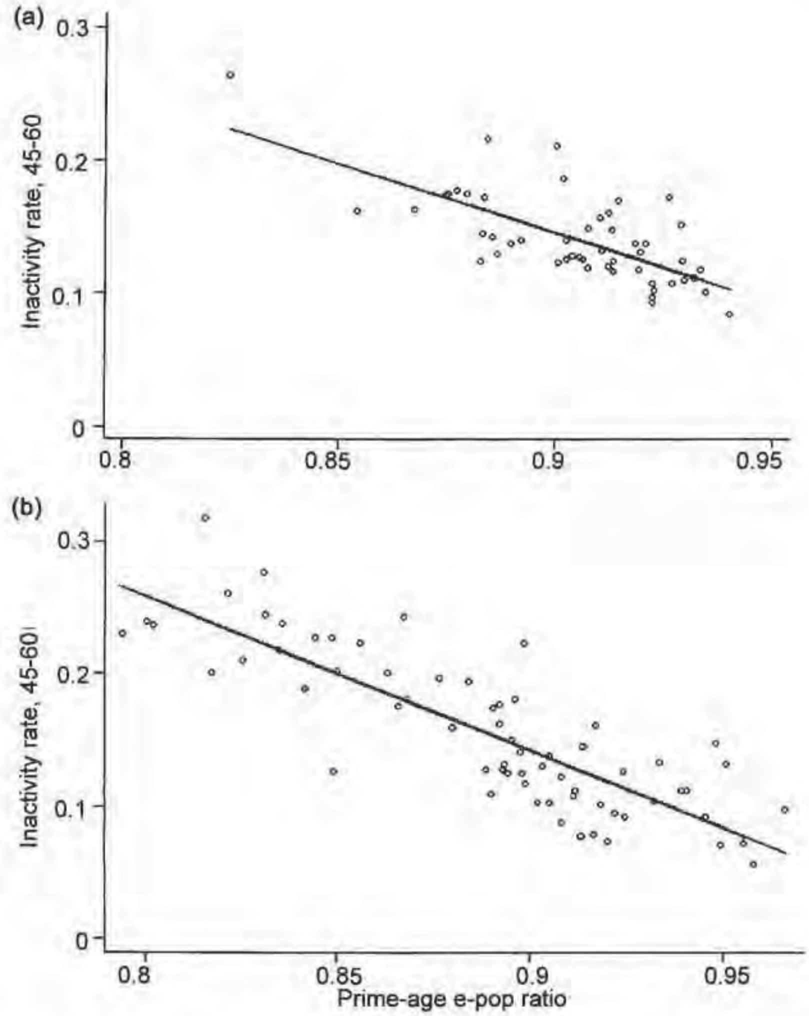

Figure 9.2 The effect of the labor market on inactivity rates of older men. (a) US states. (b) UK counties.

Notes. The US data are from the monthly CPS files for January 1998 to December 2000. Each observation is an individual state (Alaska excluded). The prime-age employment/population (e-pop) ratio is for men aged 25–44. The UK data are from the LFS November 1998 to October 1999 inclusive. Each observation is an individual county. The prime-age employment/population ratio is for men aged 25–44.

TABLE 9.2
The Impact of the Labor Market on Job Search Intensity of the Unemployed

Country	*Prime-Age Employment/ Population Ratio*	*Personal Controls*	*Region Effects*	*Time Effects*	*Number of Observations*
US	0.86 (0.12)	No	No	No	176160
US	0.72 (0.12)	Yes	No	No	176160
US	1.03 (0.21)	Yes	Yes	No	176160
US	0.49 (0.11)	Yes	No	Yes	176160
US	0.23 (0.20)	Yes	Yes	Yes	176160
UK	0.87 (0.44)	No	No	No	179979
UK	0.67 (0.30)	Yes	No	No	179979
UK	0.97 (0.53)	Yes	Yes	No	179979
UK	0.66 (0.30)	Yes	No	Yes	179979
UK	−0.89 (0.39)	Yes	Yes	Yes	179979

Notes.

1. The US data are from the basic monthly CPS for January 1994 to December 2000. The UK data are from the LFS for March 1992 to November 2000. The sample is those not in employment who report they have looked for work in the past four weeks.
2. The standard errors for the coefficient on the employment/population ratio are computed assuming clustering on region–time interactions.
3. The personal controls are gender, race, age, age squared, and four education dummies. The regional dummies are by state for the United States and by region for the United Kingdom. The time dummies are the month for the United States and the year for the United Kingdom.

market encourages job search. However, the inclusion of both regional and time controls makes the employment/population ratio change sign and be significant in the United Kingdom. It may be that there is little in the way of cyclical fluctuations left in the data once one has controlled for regional and time effects, but this result should induce some caution in the conclusions drawn.

9.3 The Job Search of the Employed

This section briefly discusses the determinants of job search among the employed. The employed will choose their search intensity to satisfy (9.5). A very clear prediction from this model is that search intensity should be decreasing in the wage. The intuition for this is straightforward: the higher the wage, the less the potential gain from job search as there are fewer higher wage jobs to find.

The US CPS does not regularly ask questions about job search among the employed but, in the Contingent Worker and Alternative Employment Supplement (conducted every two years starting in 1995), workers are asked whether they have looked for alternative employment in the past three months. In the UK LFS, questions are always asked about job search among the employed: whether they are currently looking for a different job and, in addition, questions on the search methods they use.

In the United States, approximately 6% of employed workers report having looked for another job in the past three months: in the United Kingdom approximately 7% say they are currently looking for another job and these report using an average of 3.3 search methods. Table 9.3 examines the determinants of these variables.

The first two columns present estimates for the United States of a model where the dependent variable is a binary variable taking the value 1 if the worker has looked for another job in the past three months. Separate estimates for men and women are presented as the equations are rather different. But, for both men and women the probability of looking for another job is negatively related to the wage as predicted by the model although the effect is stronger for men. More educated workers are more likely to be looking for another job, married women less likely. This last result gives extra support to the argument of chapter 7 that part of the gender gap is the result on constraints on women that prevent them from seeking out and exploiting job opportunities that may arise.

The third and fourth columns estimate a similar equation for the United Kingdom where the dependent variable is now whether the respondent is currently looking for another job. The results are very similar. Ceteris paribus, a higher wage is associated with a lower probability of looking for another job. Married women are again less likely to be looking for alternative employment and there is a very large effect from the presence of children on female job search. As the UK data have a longer time series, the regional employment/population ratio was also included as a regressor: however, its effect is not significant.

The final two columns look at another dependent variable to measure search intensity: the number of search methods used (this is only available for the United Kingdom). Those who reported that they were not looking for another job are assumed to have used no search methods. The same qualitative results are found as when the dependent variable was a 0/1 dummy: in particular the wage is negatively related to the number of search methods used.

This section has shown how, other things being equal, those in better-paying jobs are less likely to be looking for an alternative job. This is in line with the view that frictions in the labor market lead to the existence of wage dispersion for identical workers.

TABLE 9.3
The Determinants of On-the-Job Search

Dependent Variable	*Looking for Another Job*				*Number of Search Methods*	
	US Men	*US Women*	*UK Men*	*UK Women*	*UK Men*	*UK Women*
Log (wage)	−0.019	−0.016	−0.032	−0.014	−0.464	−0.273
	(0.007)	(0.007)	(0.001)	(0.001)	(0.014)	(0.021)
Married	−0.009	−0.033	0.0005	−0.0274	−0.032	−0.438
	(0.0067)	(0.006)	(0.0018)	(0.0014)	(0.029)	(0.023)
Child	−0.0084	−0.0042	−0.0034	−0.0077	−0.057	−0.162
	(0.0063)	(0.0059)	(0.0014)	(0.0013)	(0.021)	(0.025)
Black	0.003	−0.0047	0.016	0.026	0.325	0.529
	(0.008)	(0.0067)	(0.008)	(0.005)	(0.088)	(0.059)
Asian (UK only)			0.012	0.004	0.183	0.155
			(0.004)	(0.004)	(0.060)	(0.070)
Education level 1	0.080	0.041	0.052	0.040	0.743	0.632
	(0.022)	(0.017)	(0.003)	(0.003)	(0.034)	(0.041)
Education level 2	0.045	0.022	0.018	0.020	0.306	0.384
	(0.013)	(0.011)	(0.002)	(0.003)	(0.033)	(0.041)
Education level 3	0.012	0.001	0.014	0.012	0.230	0.255
	(0.010)	(0.001)	(0.002)	(0.002)	(0.035)	(0.025)
Experience/10	0.013	0.0021	0.027	0.013	0.373	0.124
	(0.010)	(0.008)	(0.002)	(0.002)	(0.031)	(0.034)

Experience/10 squared	−0.0064	−0.0022	−0.0076	−0.0051	−0.108	−0.070
	(0.0025)	(0.0021)	(0.0004)	(0.0004)	(0.006)	(0.007)
Tenure/10	−0.027	−0.032	−0.068	−0.055	−1.382	−1.231
	(0.006)	(0.006)	(0.002)	(0.003)	(0.047)	(0.059)
Tenure/10 squared	0.0016	0.0011	0.011	0.0087	0.227	0.188
	(0.0009)	(0.0004)	(0.001)	(0.0010)	(0.015)	(0.023)
Regional employment/ population ratio			−0.036	−0.047	−0.243	0.586
			(0.037)	(0.037)	(0.815)	(0.732)
Number of observations	5133	5490	167390	173858	167398	173866
Mean of dependent variable	0.062	0.057	0.076	0.065	0.26	0.22
Pseudo R^2	0.092	0.089	0.069	0.057	–	–

Notes.

1. The US data are from the Contingent Workers and Alternative Employment Supplement to the CPS in February 1997 and 1999. The UK data are from the LFS for the period March 1993 to December 2000. Education level 1 is a college degree in both countries, education level 2 is some college (US) or "A" level (UK), education level 3 is high school graduate (US) or GSCE level (UK) and the omitted education category is a high school dropout (US) or someone with no formal educational qualifications (UK). State and year dummies are also included for the United States and month, year, and region dummies are also included for the United Kingdom. For the United Kingdom, the standard errors are clustered on year–region interactions to correct the standard errors on the regional employment/population ratio: this variable is not included for the United States because of the small number of observations.
2. When the dependent variable is "looking for another job," a probit model is estimated. When the dependent variable is "number of search methods," a Poisson model is estimated.

9.4 Quits

The model of transitions from employment to non-employment used so far is extremely rudimentary as it assumes that these transitions occur at a rate that is beyond the control of workers and employers. We have already seen in section 4.5 that these transitions are often as sensitive to the wage as the job-to-job mobility rate so this model is obviously not adequate. In popular discussion, a distinction is often made between "voluntary" (what we will call quits) and "involuntary" (what we will call lay-offs) transitions to non-employment. We have already used the idea that lay-offs are involuntary when discussing the earnings losses of displaced workers in chapter 6.

But many labor economists argue that the distinction between a quit and a lay-off is not a meaningful one (see, e.g., McLaughlin 1991). A match between a worker and an employer will be dissolved when there is no surplus left, that is, the reservation wage of the worker exceeds the marginal product. When this happens, it is irrelevant whether the final wage is above the reservation wage of the worker (so there will be a lay-off) or below the marginal product (so there will be a quit). The distinction between quits and lay-offs does become meaningful if there is some form of wage rigidity. For example, Hall and Lazear (1984) considered the case where wages are fixed ex ante and not altered in response to shocks to the productivity of workers or the value of leisure. As there is evidence that such wage rigidity does exist (see chapter 5), the distinction between quits and lay-offs is a meaningful one. This section considers separations initiated by workers: the lay-off decision is considered later in section 10.6.

A natural approach to the quit decision is to assume that quits occur when the reservation wage of a worker rises above the offered wage. In chapter 5 we argued that wages are not likely to be very responsive to the reservation wage because this is plausibly the private information of the worker: this gives us the wage rigidity necessary for a theory of quits. To have an interesting model of the quit decision one obviously needs some variation in reservation wages as, for the worker to have ever started the job, their reservation wage must initially have been below the wage.

Take the model embodied in the value functions (9.2) and (9.4) and slightly modify it. First, to simplify matters, assume that job offer arrival rates are exogenous but might differ for the employed (denoted by λ_e) and the non-employed (denoted by λ_u). Secondly (and this is the crucial innovation), assume that at a rate, κ, there is a shock to the value of leisure, b, and that, when this happens, the new value of leisure has a distribution function $H(b)$. We still assume that workers face a lay-off rate δ_u that is

beyond their control. The reservation wage and separation rate are then given by the following proposition.

Proposition 9.2

1. *The quit rate to non-employment is decreasing in the wage.*

2. *The reservation wage, $r(b)$, satisfies*

$$r(b) + (\lambda_e - \lambda_u) \int_{r(b)} \frac{[1 - F(x)]dx}{\delta + \lambda_e[1 - F(x)] + \kappa[1 - H(\rho(x))]}$$

$$+ \kappa \int^{b} \frac{H(\beta)d\beta}{\delta + \lambda_u[1 - F(r(\beta))] + \kappa} = b \qquad (9.7)$$

Proof. See Appendix 9.

The first part of the proposition is easy to understand. The higher the wage, the lower is the chance that the shock to b will result in a reservation wage above w. Hence, the quit rate will be lower as we found in table 4.7.

The second part of the proposition shows the way that the reservation wage rule needs to be modified if the reservation wage is stochastic. In fact, in the case this chapter has argued is the most plausible ($\lambda_u \geq \lambda_e$), the impact of potential change in the value of leisure is always to lower the reservation wage below what it would otherwise have been. This means that a worker who thinks that their value of b might change in the future will be prepared to take a job that someone with the same current value of b but no prospect of future change would not. For example if $\lambda_u = \lambda_e$ then, in the absence of variation in the value of leisure, the reservation wage is equal to b. But, once there is potential variation in b, the reservation wage will be less than b. The intuition is simple. The option to quit employment for non-employment is always present but the ability to get work is not. Consequently, workers will be prepared to take jobs they might not want now in the hope that b will fall in the future and they will then get utility gains from being in work.

This section has shown how one can provide a more adequate model of quits. However, a model with endogenous quits is messy to work with and we will not use it in what follows. The most important conclusion is that we would expect quits to non-employment to be negatively related to the wage as is observed in the data.

9.5 Involuntary Unemployment

Several sorts of unemployment have appeared in previous chapters. There is unemployment caused by the fact that it takes time for an unemployed worker to find an employer: this is frictional unemployment as conventionally understood. Secondly, if there is heterogeneity in reservation wages (e.g., the model of section 3.4) some potentially productive matches may not be consummated because the wage offered is below the reservation wage of the worker. The resulting unemployment is often described as voluntary unemployment (as it is the worker who vetoes the match) and we will follow this terminology here.[7] Although it is sometimes useful from the conceptual point of view to divide unemployment into these two components, one should not think of them as being determined independently. For example, the extent of frictional unemployment depends on the reservation wage that will be determined by the wage offer distribution which will itself determine the extent of voluntary unemployment.

But, the types of unemployment we have described are a poor description of unemployment as experienced by many of the unemployed. In the models used so far, a worker could get a job offer (at some wage) from any employer simply by approaching them. If they remain unemployed that is only because this wage offer is below their reservation wage. But the predicament of the unemployed is often not so much the problem of finding a job paying an acceptable wage as a problem of employers refusing to offer any employment at all. This is what is often called involuntary unemployment. Involuntary unemployment is said to exist when unemployed workers strictly prefer employment to unemployment at prevailing wages (and identical workers are in employment at those or better terms) but cannot obtain employment (see, e.g., the definition offered by Taylor 1987). In a static model of a perfectly competitive labor market, this corresponds to employment being on the labor demand curve and off the labor supply curve. The obvious question to answer in this case is why wages do not fall to clear the labor market. One explanation is that institutions like minimum wages or unions prevent this wage adjustment but many economists feel that involuntary unemployment exists even in unfettered labor markets and there has been a considerable amount of

[7] This terminology does have some unfortunate connotations. For example, inefficient voluntary unemployment is often thought of as being "caused" by excessive levels of welfare benefits. But, in the model here it is "caused" by wages being low relative to reservation wages and this could equally be due to wages being too low or reservation wages too high. The traditional one-sided view of the problem is the result of implicitly using a competitive model in which one thinks of the wage as being the marginal product independent of the level of benefits. But, in a monopsonistic labor market, this is not the case.

research into the reasons why this might be the case, a branch of research that has been grouped together into efficiency wage theory.

At first sight, one might think that efficiency wages and monopsony are incompatible as the simplest efficiency wage model has an equilibrium on the labor demand curve but off the labor supply curve while the simplest monopsony model has an equilibrium on the labor supply curve but off the labor demand curve. However, as we shall see, one can reconcile the existence of involuntary unemployment with an upward-sloping supply curve of labor to the employer.

One immediate problem is that the usual definition of involuntary unemployment is not well suited to a labor market with frictions. The definition is typically applied in models where there are well-defined labor demand and supply curves, and there is a single wage in the market. Our first task is to suggest a definition more suited to our purposes.

Definition. *Involuntary unemployment exists whenever, given a match between an unemployed worker and a firm, the match is not consummated even though the worker wants the job at the offered terms and there are equally productive workers employed in this or identical firms at terms that the worker would also accept.*

Several parts of this deserve discussion. First, by considering involuntary unemployment to exist only when a match fails to be consummated and not the absence of a match, we avoid defining as involuntary unemployment a situation in which the worker knows there is a single high wage job in the economy but does not know where it is. The requirement that the worker wants the match on the offered terms corresponds to the fact that the worker would like to be employed at the offered contract. The condition that other equally productive workers are employed at these or better terms ensures that the worker could "reasonably" aspire to getting the job.

It should be apparent that involuntary unemployment on this definition does not exist in the basic models of the labor market we have used so far. When a match is made, the employer will always want to employ the worker if the wage is below the marginal product and the worker is prepared to do the job. The employer would not be prepared to employ the worker at a wage above his/her marginal product but no worker would be obtaining such a wage so that is not classed as involuntary unemployment.

But, as the following section shows, it is possible for monopsony and involuntary unemployment to be compatible. This is shown using the most common models of involuntary unemployment, efficiency wage models.

9.6 Efficiency Wages and Monopsony

Akerlof and Yellen (1986) describe four main types of efficiency wage models: the shirking model, the turnover model, the adverse selection model, and the fairness model. We consider each of these in turn.

9.6.1 *The Shirking Model*

As presented by Shapiro and Stiglitz (1984), the idea of the shirking model is that monitoring of workers is less then perfect. To ensure that workers put in an appropriate level of effort, workers who are caught shirking must be punished in some way. It is assumed that the only form of feasible wage contract is a fixed wage if employed, so that the worst punishment that can be inflicted on a worker is to be fired. But, if workers are as well off when unemployed as when employed, they will not care whether they keep or lose their jobs so that they will always shirk. So, employers must always offer workers a contract such that they are strictly better off in work (and not shirking) than out of work. But, this will mean that some unemployed workers cannot obtain employment even though they would be prepared to work for lower wages than currently employed workers. So, equilibrium is off the labor supply curve and unemployment is involuntary.[8] Now consider how these ideas can be put into the models we have used.

Assume that workers are heterogeneous in b, the utility received when non-employed and that the distribution function of b is given by $H(b)$. A worker employed in a job paying wage w and not shirking has utility $(w - e)$ while a shirking worker has utility w. Workers who do not shirk have a marginal product equal to p but workers who shirk produce nothing. Job offers arrive at a rate λ for both employed and unemployed workers (this makes the analytics simpler without losing any of the features of the model that we want to emphasize here). These assumptions imply that the reservation wage of a worker who does not shirk is $(b + e)$. Workers also exit the labor force at an exogenously given rate δ.

[8] It should be noted that there is a debate about whether the results of this model are a product of an arbitrary restriction in the form of employment contracts (see Carmichael 1985; Macleod and Malcomson 1989). This is essentially a debate about whether workers can be made to post bonds that are lost if their performance is inadequate. This is an issue that we have already considered when we discussed the ways in which the employer might become a discriminating monopsonist in chapter 5. The upshot of our discussion there was that, while there are reasons to think that employers will try to find ways to use more sophisticated contracts, there are also good reasons to think that their ability to eliminate all the problems is less than perfect. In this case, the insights of the simple model will still be relevant.

Assume that employers set a wage for the job and, while they can observe b, they do not attempt to fine tune their wage offers to the individual characteristics of job applicants.[9] We also follow Shapiro and Stiglitz (1984) in assuming that the only available form of labor contract is a fixed wage with shirkers being fired. Together, these two assumptions mean that employers turn away workers who they think will shirk at the going wage. We assume there is imperfect monitoring of workers; in particular that the incidence of monitoring is a Poisson process with arrival rate θ. Workers who are caught shirking are assumed to be fired.

The following proposition shows that the minimum wage at which an employer is prepared to hire a worker is strictly above their reservation wage.

Proposition 9.3. *An employer will refuse to hire some workers who want employment at the offered wage. The minimum wage, $w(b)$, that must be paid to a worker to prevent them from shirking solves*

$$w(b) = b + e + \frac{\delta + \lambda[1 - F(w(b))]}{\theta} e \qquad (9.8)$$

Proof. See Appendix 9.

An employer will only want to employ the worker if, given b, the wage offered is above $w(b)$ as given by (9.8). As $w(b) > (b + e)$, the reservation wage of workers, (9.8) implies that the employer will sometimes be turning away workers who want the job. The gap between $w(b)$ and the reservation wage is larger if monitoring is less effective (θ is smaller), and the separation rate $[\delta + \lambda(1 - F(w(b)))]$ is larger. The intuition for the last result is that the worker is more worried about losing a job that is expected to last a long time. Using the definition provided, some workers are involuntarily unemployed as long as some identical worker (i.e., with the same b) is in employment in the economy at a wage at which this worker would also like work. This will be the case if there are any workers with the same level of b in employment in the economy. Those in employment in this model will be receiving higher wages.

[9] Some discussion of this is in order. In chapter 5 we argued that wages were not likely to depend on b because it was likely to be private information of the worker. But if this was the case in the model here, employers would be unable to tell which workers will shirk and which will not. So, we are implicitly assuming that firms set a single wage for the job, something for which we have argued in chapter 5 that there is considerable empirical evidence. One could obtain the results reported here as long as employers have some information on b which is not incorporated into the wage offer.

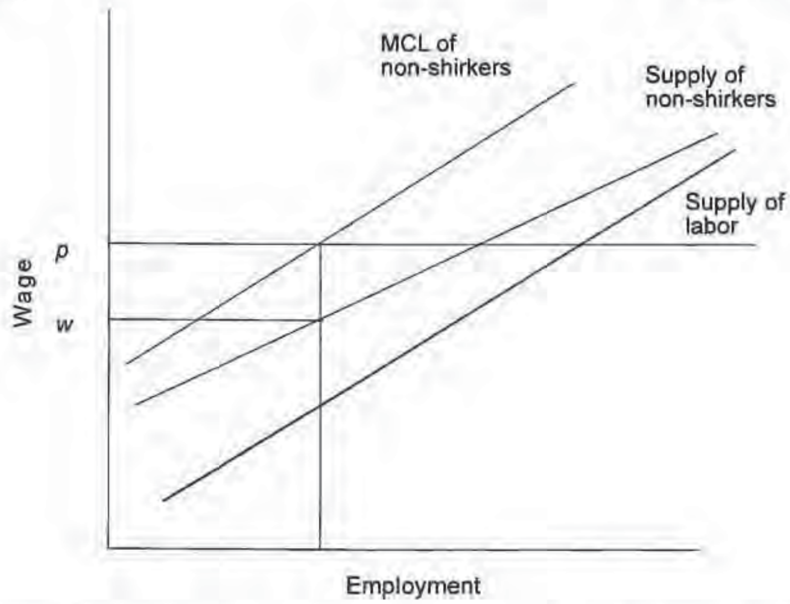

Figure 9.3 Involuntary unemployment and monopsony in the shirking model.

It is simple to provide a diagram to explain how one can reconcile this involuntary unemployment and monopsony in the shirking model. Because of labor market frictions and heterogeneous workers, the supply curve of labor to the firm is not perfectly elastic and is as drawn in figure 9.3. But, the employer knows that some of the workers who want to work in the firm will shirk so that the supply of non-shirkers lies somewhere to the left of the labor supply curve as also drawn in figure 9.3. It is the gap between the two lines that gives the existence of involuntary unemployment. The firm, like any good monopsonist, will choose a level of employment where the marginal cost of labor derived from the supply curve of non-shirkers is equal to the marginal revenue product of labor as drawn in figure 9.3. In figure 9.3, this results in the choice of wage, w.

In this section we have shown how one can introduce the ideas of Shapiro and Stiglitz (1984) into our model of the monopsonistic labor market.[10] The result is a model in which there are both features of monopsony and involuntary unemployment: the insights of these models are not incompatible. Let us now consider other versions of the efficiency wage model.

[10] See Manning (1995) for a way to do this in a general equilibrium framework.

9.6.2 *The Turnover Model*

In Salop (1979), the basic idea is that the labor turnover rate depends on the wage and that the presence of turnover costs then means that an employer may not be prepared to cut wages to employ workers who cannot get work. As the idea that turnover depends on the wage is central to this book, his model can be expressed simply using our existing terminology.

Assume that the firm incurs total turnover costs of $T(R)$ where R is the recruitment rate.[11] If $s(w)$ is the separation rate then employment, N, will be given by $N = (R/s)$. Assume that revenue is given by $Y(N)$. Profits in a steady state can be written as

$$\pi = Y(N) - wN - T(s(w)N) \tag{9.9}$$

The employer chooses the wage and the level of employment to maximize (9.9) subject to the constraint that

$$N \leq \frac{R(w)}{s(w)} \tag{9.10}$$

where $R(w)$ is the flow of recruits to the firm. Involuntary unemployment occurs whenever the constraint in (9.10) does not bind, that is, the firm does not want to hire all the recruits attracted to the firm.

There are some cases in which involuntary unemployment cannot occur. If the production function has constant returns to scale and the costs of turnover are linear in the number of recruits, then the profit function of (9.9) will be linear in N and, assuming the employer wants to hire any workers at all, it will want to hire all the workers it can. But, if there are decreasing returns to scale or increasing marginal turnover costs, it is quite possible that the employer will not want the constraint in (9.10) to bind.[12]

However, there is something a little unsatisfactory about this as a model of involuntary unemployment. The supply of labor to the firm as written in (9.10) assumes that the employer only has a single instrument, the wage, to influence the flow of recruits to it. If, as in the generalized model of monopsony of section 2.3, the employer can also choose the

[11] There are other specifications of the turnover cost function that might be plausible. For example, one might assume that total turnover costs can be written as a function $T(R, N)$ where $T(\cdot)$ is homogeneous of degree one in its arguments. As $sN = R$ in equilibrium this amounts to having a turnover cost function of the form $RT(s)$ which, if s depends on w implies a per worker turnover cost that is a declining function of w.

[12] Manning (1995) shows that, in the case where $T(R)$ is a convex function of R, a minimum wage that just binds must raise employment even if there is involuntary unemployment.